Experimental Data Analysis of Toronto Police Service Strip Search Rates

Ragave Vicknarajah (1003959610)

Kailai Yang (1005536939)

INF2178: Experimental Design for Data Science

Dr. Shion Guha

April 15, 2023

Introduction

Police have the power to conduct strip searches to ensure safety and security when they suspect a person may be in possession of a weapon or other contraband. They can also be used to keep illegal items like narcotics or weapons out of correctional facilities. (Babbar, 2021) However, concerns have been raised about the use of strip searches and the potential for abuse or infringement of civil liberties. They are a highly invasive procedure that can be traumatic and degrading to the person being searched. As a result, there is a lot of debate regarding whether or not to conduct strip searches, with concerns like privacy, human rights, and abuse potential being raised. Strip searches can be a serious violation of privacy, especially for vulnerable populations such as minors. To investigate the impartiality of Toronto police officers' requests for strip searches during stops of various minority groups, we will examine the demographics of those asked to strip search from arrest. We hypothesize that the Toronto police maintain consistent fairness in requesting strip searches for various minority suspects during the 2020-2021. We will use the dataset containing information related to all arrests and strip searches data publicly available on the Toronto Police Service Public Safety Data Portal during the 2020-2021 years for this project.

Literature Review

A strip search is a sort of search in which a person's clothing is partially or completely removed in order to look for contraband or other objects. The use of strip searches is debatable since they may be viewed as an invasion of privacy and dignity, especially if they are performed in a way that is humiliating or public. (Babbar, 2021) This can be especially traumatic for vulnerable populations such as minors, the elderly, and survivors of sexual violence. (Psutka & Sheehy, 2016).

In addition, there is a growing awareness of systemic bias and discrimination in law enforcement and the criminal justice system. Racial and ethnic discrepancies can be found at every level of the criminal justice process, from arrest to sentence, according to a growing corpus of studies. For instance, research has shown that compared to white defendants, black and Hispanic defendants are more likely to face more serious charges and to be sentenced to lengthier terms of imprisonment.(Smith et al., 2020) Research shows that people of color are disproportionately subjected to strip searches, even when they are not suspected of carrying contraband or weapons. These searches can have a traumatic impact on individuals who have already experienced systemic discrimination and racial profiling. (Alexander, 2012)

Strip searches are often a topic of concern when it comes to gender issues because of the specific issues that arise when strip searches are conducted on people of different genders. For example, women may be more likely to experience trauma and emotional distress as a result of strip searches due to prior experiences of sexual violence and objectification. (Lemke, 2022) Because strip searches can be a traumatic and invasive experience, especially for minors who may find greater physical and psychological effects. According to the report by the National Prison Rape Elimination Commission (NPREC) found that over half of the juvenile facilities surveyed reported conducting strip searches on all youth entering the facility, regardless of individualized suspicion. (2009) There is a need to strengthen protections for vulnerable populations, including clear guidance for minors on when and how to conduct strip searches. Therefore, police officers should be more cautious when making strip searches on suspects, especially when dealing with some vulnerable groups.

Dataset Description

The dataset we used in this project shows information related to arrests and strip searches conducted by police officers, as well as indicators of whether a person was booked at a police station within 24 hours following the arrest event. It could be used to study patterns of arrest and strip search behavior among police officers, as well as the relationship between strip searches and booking rates. This dataset has 65,276 Records in which the basic information such as perceived race, gender, and age group of each captured person is recorded during the 2020-2021 years. The dataset - Arrests and Strip Searches (RBDC-ARR-TBL-001) - is available on Toronto Police Service Public Safety Data Portal and can be found through the following link: https://data.torontopolice.on.ca/datasets/TorontoPS::arrests-and-strip-searches-rbdc-arr-tbl-001/about. The dataset includes the following information: Arrest Year, Arrest Month, EventID, ArrestID, PersonID, Perceived Race, Sex, Age group (at arrest), Youth at arrest (under 18 years), ArrestLocDiv, StripSearch, Booked, Occurrence Category, Actions at Arrest - Concealed items, Actions at Arrest - Combative, violent or spitter/biter, actions at arrest - resisted, defensive or escape risk, Actions at arrest - Mental instability or possibly suicidal, Actions at arrest - Assaulted officer, Actions at arrest - Cooperative, SearchReason-CauseInjury, SearchReason-AssistEscape, SearchReason-PossessWeapons, SearchReason-PossessEvidence, ItemsFound. The strip search indicator is a binary variable indicating whether a person was subject to a search that involved the removal of some or all clothing and a visual inspection of the body. The booking status is a binary variable indicating whether a person was booked within 24 hours following the arrest event. Some records may indicate that a person was strip searched, but the data does not indicate a booking; in those cases, it is assumed that a booking took place. The age of the person arrested and/or strip searched is the age they provided to the arresting officer at the time of arrest.

Research Objective and Questions

The objective of our project is to analyze the arrest and strip search dataset provided by the Toronto Police. Our focus is to examine whether individuals are treated equally when it comes to being asked to undergo a strip search, with a specific focus on variables such as race, age, sex, and youth status. Through our analysis, we aim to identify any disparities or biases that may exist in the strip search practices of the Toronto Police, and to provide recommendations for any necessary changes that could help to ensure that all individuals are treated fairly and with respect during interactions with law enforcement. By shedding light on potential issues in strip search practices, our research aims to contribute to broader efforts towards promoting equity and justice within the criminal justice system.

We propose some research questions that can help guide our analysis of the Toronto Police arrest and strip search dataset to determine whether certain demographic groups are being treated differently when it comes to strip searches. Based on our literature review and preliminary analysis of the data set, we have identified the possibility of bias and differential treatment of minorities in the current justice system. In order to further explore this issue, we have formulated the following research questions.

- RQ: Is there a significant association between an individual's demographic characteristics (race, sex, age group) and the probability of being subjected to a strip search, while accounting for temporal changes in strip search rates?
    - RQ1: Are individuals of certain races, or non-white people more likely to be subjected to strip searches by the Toronto Police based on the arrest year?
    - RQ2: Is there a significant difference in the frequency of strip searches for male and females in Toronto Police arrest cases?

We believe an in-depth study of this dataset will help us explore these research questions and help us understand the reasonableness of the Toronto police in asking suspects to strip search.

Descriptive Statistics

Data cleaning

The original dataset provided on Toronto Police Service Public Safety Data Portal was available in a CSV file which contains a total 65, 276 Records. The entire dataset includes only categorical variables and binary variables, and does not provide any numerical, which does not allow us to use the variables in the dataset directly as dependent variables for analysis using anova test. In order to run the ANOVA and t-test on the dataset it needs to be cleaned since there are columns with N/A values and no continuous variables. To begin we loaded the dataset into Google Collab and right away checked for any columns that contained N/A values. The columns that included N/A values were the following: ArrestID, Age_group_at arrest_, Perceived_Race, and Occurance_Category. The rows that contained N/A values were dropped from the dataset. After cleaning the dataset we were left with 64,615 records and 10 columns of data.

In order to include a continuous variable we decided to transfer the categorical variables to numerical variables. Arrest year includes two levels, 2020 and 2021, which we transformed into 0 to represent 2020 and 1 to represent 2021. Perceived Race includes a total eight categories , which we transformed into 0 to represent Black, 1 to represent East/Southeast Asian, 2 to represent Indigenous, 3 to represent Latino, 4 to represent Middle-Eastern, 5 to represent South Asian, 6 to represent Unknown or Legacy and 7 to represent White. Due to the relatively small sample of non-white groups, we merged others other than white in what is collectively referred to as Non-White. In the Non-White column 0 means the person is white and 1 represents non-white people. Sex includes 3 levels, Male, Female and Unknown, which we encoded Female to 0, Male to 1 and Unknown to 2. The original datasets contain 9 age groups, which they divided 17 years and younger and 65 years and older twice. We therefore combined all those under 17 years of age, and the same for those over 65 years of age. The new merged age group contains seven groups, which we transformed into 0 to represent Aged 17 and under, 1 to represent Aged 18 to 24 years, 2 to represent Aged 25 to 34 years, 3 to represent Aged 35 to 44 years, 4 to represent Aged 45 to 54 years, 5 to represent Aged 55 to 64 years, 6 to represent Aged 65 and older.

The final dataset after cleaning up has 10 columns of data:Arrest Year, Arrest Month, Perceived Race, Sex, Age group at arrest, youth at arrest under 18 years, Non white, StripSearch, Booked, Occurrence Category and Items Found. Since this dataset uses only binary variables and our research question focuses more on strip search, we counted the number of times strip search occurred on each group and named it Search count. Because the number of groups in the sample varies, the mere fact that search count is not sufficient to account for the fairness of police in asking suspects to strip search. We counted the proportion of strip searches for further analysis, and this data was named Search count. Search rate is the proportion that the arrest people are asked to do the strip search in the total group. Search rate equals to stripe_saerch_count divided to the whole group (strip_search_count + not_Strip_search_count). The new merged dataframe after grouped includes the following columns: Arrest Year, Perceived Race, Sex, Age group at arrest, youth at arrest under 18 years, Non white, Search count and Search rate.

Data Summary

The original dataset is grouped by the Arrest Year, Perceived Race, Sex, Age group at arrest, youth at arrest under 18 years and Non white columns. After grouped the dataset we were left with 230 records and 8 columns of data. From Figure 1, we can find the mean of these 230 search count records is 31.88 and mean of search rate is 7.28%. The standard deviation is 100.318 in search count and 9.36% in search rate. The max group has the 804 search count and 55.56% search rate.

	Search_count	search_rate
count	230.000000	230.000000
mean	31.873913	7.280802
std	100.318126	9.365414
min	0.000000	0.000000
25%	0.000000	0.000000
50%	3.000000	2.679115
75%	17.750000	13.596931
max	804.000000	55.555556

Fig.1. Dataset Summary

Race

The perceived races included in the dataset are: White, Black, Unknown or Legacy, East/Southeast Asian, South Asian, Middle-Eastern, Indigenous, and Latino. According to the graph and the frequency counts (Figure 2) majority of the people included in the dataset were perceived as White compared to the other races included in the dataset. Figure 3 includes a count bar graph of the Perceived_Race category from the dataset to show the counts of whether the person got strip search.

```
White                    27407
Black                    17352
Unknown or Legacy         5002
East/Southeast Asian      4388
South Asian               3594
Middle-Eastern            3213
Indigenous                1907
Latino                    1752
Name: Perceived_Race, dtype: int64
```

Fig.2. Strip Search Counts on Each Races

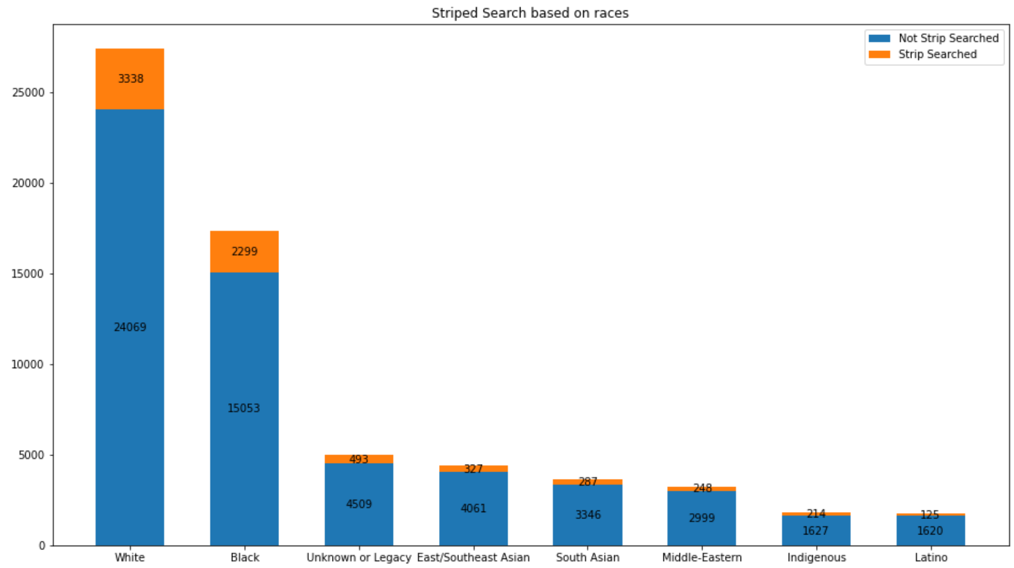

Fig.3. Strip Search Proportion on Each Races

Gender

The gender included in the dataset are: Female, Male and Unknown. Figure 4 includes the total count number in each gender group. According to the counts and bar chart (Figure 5), we can find the proportion of not strip searched and strip searched in each gender group. 6,123 Males and 1,208 females are requested to strip search. In this cleaned dataset, it only contains 9 records about unknown gender and all of them were not searched.

```
M    52106
F    12500
U        9
Name: Sex, dtype: int64
```

Fig.4. Strip Search Counts on Each Gender Group

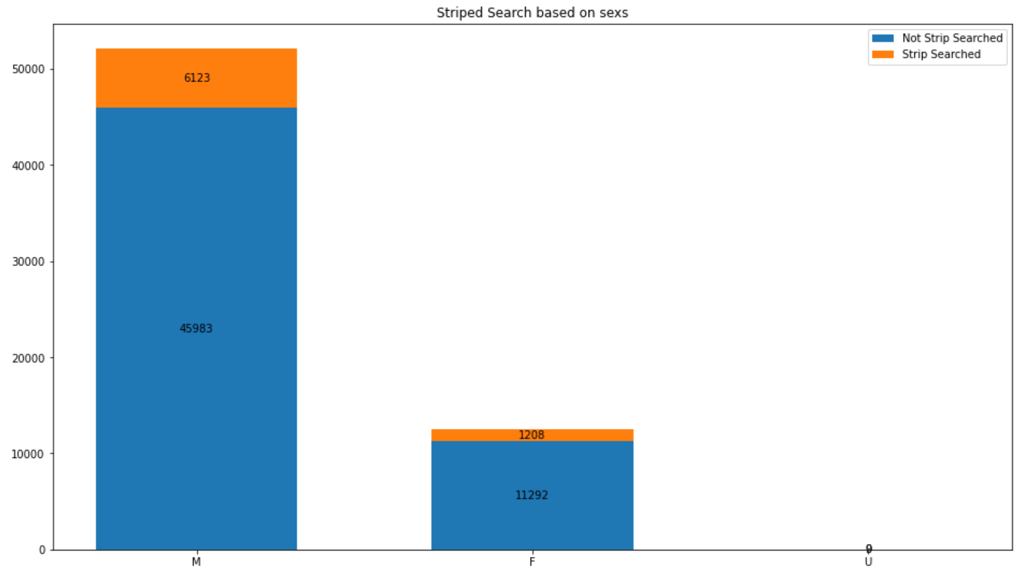

Fig.5. Strip Search Proportion on Each Gender Group

Correlation Plot

The correlation heatmap (Figure 6) provides a visualization of the strength of relationships between the numerical variables (Kumar, 2022). Using the correlation heatmap we are able to have an idea of the strength of the relationship of the variables (Kumar, 2022). A value closer to one is considered a positive correlation between the two variables, a value closer to negative one is considered a negative correlation, and if the value is zero there is no correlation between the variables (Kumar, 2022).

According to Figure 6 there are no variables that have a strong positive correlation, meaning that there are no variable relationships close to one. The most notable positive relationships are between the perceived race of the individual and age group at arrest, 0.18. However, this relationship is closer to zero than one suggesting that there is no real correlation between the variables. There is a negative correlation between age group at arrest and youth at arrest under 18 years. However, the correlation does not have much significance since the variables only have a relationship because the data collected for those variables are relatively the same.

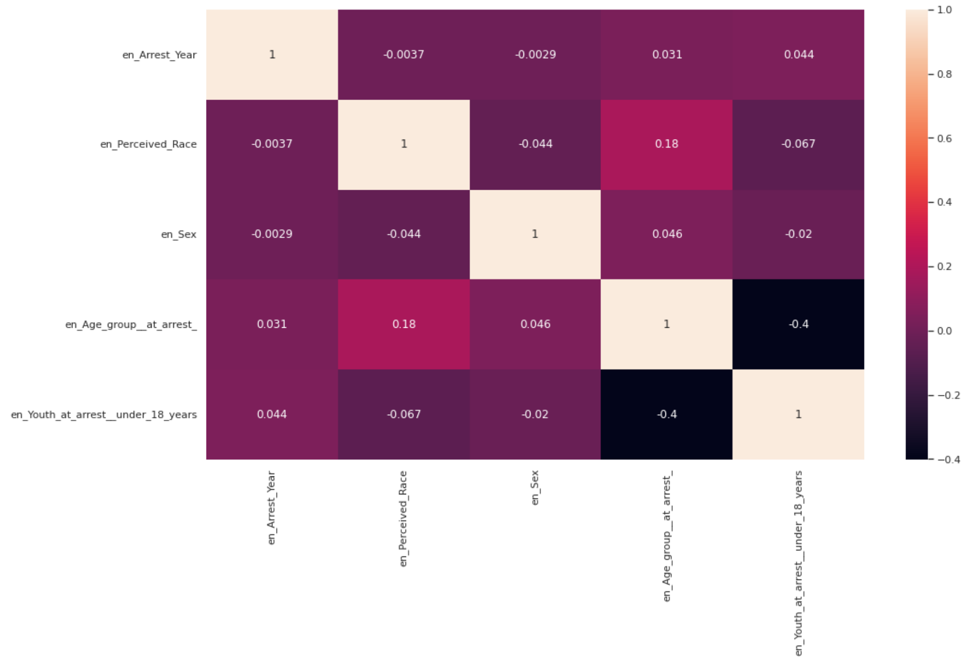

Fig.6. Correlation Plot

Methods

T-Tests

A t-test is a statistical test that compares the means of two samples (Bevans, 2022). It is used to determine if there is a significant difference between the means of two groups and how they are related (Bevans, 2022). The test uses a null hypothesis that the difference in the group means is zero and an alternative hypothesis that the difference in group means is different from zero. (Bevans, 2022) For this paper six t-tests were carried out to examine the relationship between two selected variables of our choosing from the dataset (can be found in the results section).

- T-Test 1(p-value: 0.469 > 0.05):
  Hypothesis: The arrest white people and non white people has the same search rate
  Null Hypothesis (H0): Mean search rate white = Mean search rate non white
  Alternative Hypothesis (Ha): Mean search rate white != Mean search rate non white

- T-Test 2 (p-value: F to M - 0.011; F to U - 0.055; M to U - 0.023):
  Hypothesis: Female and Male has the same search rate
  H0: Mean search rate female = Mean search rate male = mean search rate unknown
  Ha: Mean search rate female != Mean search rate male != mean search rate unknown

- T-Test 3 (p-value: 0.0.847 > 0.05):
  Hypothesis: white female and white male has the same search rate
  H0: Mean search rate white female = Mean search rate white male

Ha: Mean search rate white female != Mean search rate white male

- T-Test 4 (p-value: 0.761 > 0.05):
  Hypothesis: adult and youth has the same search rate
  H0: Mean search rate adult = Mean search rate youth
  Ha: Mean search rate adult != Mean search rate youth

- T-Test 5 (p-value: 0.653 > 0.05):
  Hypothesis: white and black has the same search rate
  H0: Mean search rate white = Mean search rate black
  Ha: Mean search rate white != Mean search rate black

- T-Test 6 (p-value: 5.289e-31 > 0.05):
  Hypothesis: In 2020 and 2021 has the same search rate
  H0: Mean search rate white = Mean search rate black
  Ha: Mean search rate white != Mean search rate black

Power Analysis

A power analysis was conducted in order to determine the appropriate sample size required to detect significant effects with adequate statistical power. Power analysis is a statistical method used to determine the smallest sample size needed for an experiment, given a required significant level, statistical power, and effect size (TIBC, n.d.; Statistics Solutions, n.d.). The main purpose of power analysis is to help the researcher determine the smallest sample size that is suitable to detect the effect of a given treatment or intervention (Statistics Solutions, n.d.).

ANCOVA

ANCOVA stands for Analysis of Covariance, which is a statistical method used to analyze the relationship between a dependent variable and one or more independent variables, while controlling for the effects of one or more covariates (Laerd, n.d.). For this research paper a one-way ANCOVA was conducted, it is a type of ANCOVA that is used when there is one categorical independent variable (with three or more groups) and one continuous dependent variable, and the effect of the independent variable on the dependent variable is assessed while controlling for the effects of one or more covariates (StatsTest, 2023). The ANCOVA will help to assess whether there are significant differences in the strip search rates between different racial groups, while controlling for potential covariates such as past strip search rates or age.

In order to run an ANCOVA, there are a series of assumptions that need to be met. ANCOVA follows similar assumptions as in ANOVA, specifically for the following: the samples should be independent (Mackenzie, 2018), the populations from which the samples are obtained must be normally distributed, and the variance of each group is equal (homogeneity of variance). In addition, ANCOVA also needs to meet the following assumptions: 1) linearity assumption, 2) homogeneity of within-group regression slopes (parallelism or non-interaction), 3) dependent variable and covariate should be measured on a continuous scale, 4) covariate should be measured without error or as little error as possible (Bedre, 2022).

Prediction Interval

Prediction intervals will be conducted in order to estimate the range of values within which future strip search rates for individuals are expected to fall, with a certain level of confidence. Prediction intervals is an estimate of an interval in which a future observation will fall, given a set of predictor variables (Lewis & Bruner, 2022). It provides a measure of reliability for the prediction of an observation and is different from a confidence interval (Lewis & Bruner, 2022).

Logistic Regression

Logistic regression is a statistical analysis method used to predict a binary outcome, such as yes or no, based on prior observations of a dataset (Lawton et al., 2022). It is a type of regression analysis that estimates the parameters of a logistic model, which is a linear combination of the predictor variables (Lawton et al., 2022). The goal of logistic regression is to find the best fitting model that describes the relationship between the predictor variables and the binary outcome variable (Thanda, 2022). For the purpose of this research paper a logistic regression was used to quantify the association between demographic characteristics (e.g., race, sex, age group) and the binary outcome of whether or not an individual was subjected to a strip search. This analysis allows us to assess the individual contribution of each demographic variable while controlling for others.

The following assumptions must be met for a logistic regression to be valid: 1) the dependent variable must be binary or dichotomous, 2) there should be no multicollinearity among the predictor variables, 3) the relationship between the predictor variables and the log odds of the outcome variables should be linear, 4) there should be no outliers or influential observations that can affect the model fit, 5) the sample size should be large enough to ensure stable estimates of the model parameters (Statistics Solutions, n.d.).

Results

Power Analysis

Male vs. Female

The effect size retrieved using Cohen's D for strip search was 0.07. The statistical measure Cohen's D was used because the purpose of the analysis is to test the difference between male and female strip searches. The value from Cohen's D statistical measure was too big to have a medium effect (d = 0.5), and too small to have a large effect (d = 0.8) (Cohen, 1988). The significance level that will be used throughout the power analysis will be 5% or 0.05 and a power level of 0.80. The sample size needed for Female_StripSearch was 2246, and for Male_StripSearch it was 9364. The power analysis for male versus female strip search resulted in a statistical power value of 0.877 (rounded to 0.9). A power value of 0.9 indicates that there is a 90% chance of detecting a difference between our two groups of interest, male and female.

The plot in Figure 7 demonstrates the power curves for different effect sizes (including a representative size of 0.07 and additional effect sizes of 0.2, 0.5, 0.8). The x-axis represents the sample size, and the y-axis represents the power. Each curve represents the relationship between the sample size and the achieved power for a specific effect size. The vertical red dashed line represents the representative sample size for Female_StripSearch (2246), which is the sample size needed to achieve the desired power level for that group. The vertical blue dashed line represents the representative sample size for Male_StripSearch (9364), which is the sample size needed to achieve the desired power level for that group. The plot provides a visual representation on how statistical power changes with varying sample sizes, especially for different effect sizes. It also shows the calculated sample sizes for the two groups in the context of power curves.

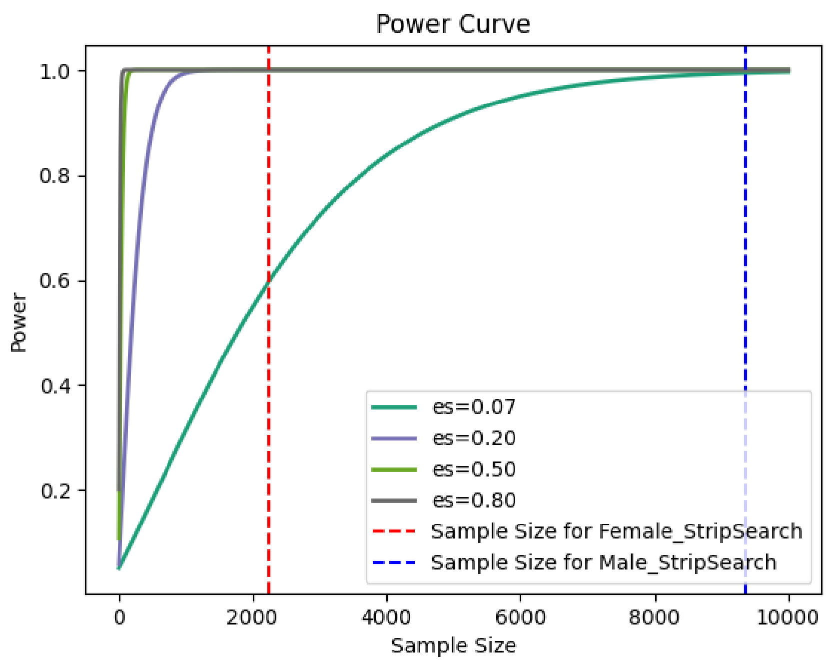

Figure 7: Power curve from the analysis of female vs. male strip searches

Based on the plot from Figure 7, we can infer that an increase in the sample/effect size leads to an increase in power. In other words, the bigger the sample, the higher the power, keeping other parameters constant.

White vs. Non-White

The effect size retrieved using Cohen's D for strip search was 0.04. The statistical measure Cohen's D was used because the purpose of the analysis is to test the difference between male and female strip searches. The value from Cohen's D statistical measure was too small to have a medium effect (d = 0.5) (Cohen, 1988). The significance level that will be used throughout the power analysis will be 5% or 0.05 and a power level of 0.80. The sample size needed for Non_White_StripSearch was 8876, and for White_StripSearch it was 6538. The power analysis for male versus female strip search resulted in a statistical power value of 0.990. A power value of 0.9 indicates that there is a 90% chance of detecting a difference between our two groups of interest, male and female.

The plot in Figure 8 shows the power curves for different effect sizes (including a representative effect size of 0.046 and additional effect sizes of 0.2, 0.5, 0.8). The x-axis represents the sample size, and the y-axis represents the power. Each curve represents the relationship between the sample size and the achieved power for a specific effect size. The vertical red dashed line represents the representative sample size for Non_White_StripSearch (8876), which is the sample size needed to achieve the desired power level for that group. The vertical blue dashed line represents the representative sample size for White_StripSearch (6538), which is the sample size needed to achieve the desired power level for that group. The plot provides a visual representation of how statistical power changes with varying sample sizes, especially for different effect sizes. It also shows the calculated sample sizes for the two groups (Non_White_StripSearch and White_StripSearch) in the context of the power curves.

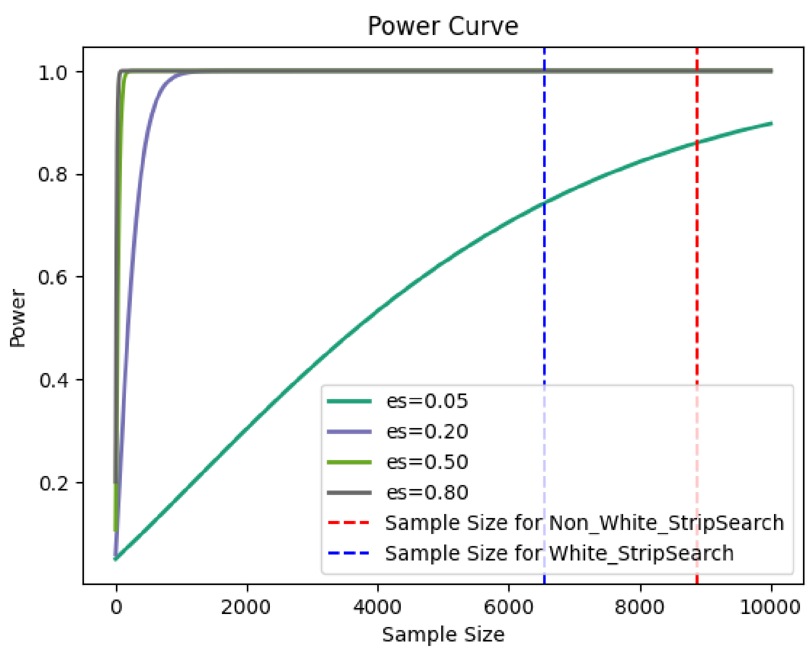

Figure 8: Power curve from the analysis of white vs. non-white strip searches

ANCOVA

The purpose of this research paper is to identify if there is a significant association between an individual's demographic characteristics (race, sex, age group) and the probability of being subjected to a strip search, while accounting for temporal changes in strip search rates? In order to test this hypothesis two one-way ANCOVA statistical tests were conducted. Additionally, to run the ANCOVA there needs to be one independent variable, one dependent variable, and the covariate (or variable that moderates the impact of the independent on the dependent variable).

One-way ANCOVA #1

The null hypothesis for the first one-way ANCOVA is there is no significant difference in the mean StripSearch_Count between the different levels of the categorical variable Perceived_Race after controlling for the continuous covariate Age_group_numeric. The alternative hypothesis is that there is a significant difference in the means of the variables that were identified in the null hypothesis.

The results of the first one-way ANCOVA can be found in Table 1 and 2. In Table 1, the p-value for the Perceived_Race is 0.000009, which is statistically significant at a 0.05 significance level. This means that we reject the null hypothesis and conclude that there is a significant difference in the mean StripSearch_Count between the different levels of Perceived_Race after controlling for Age_group_numeric. The p-value for the Age_group_numeric variable is approximately 0.072844, which is not statistically significant at a 0.05 significance level. This means that we fail to reject the null hypothesis for this covariate, and we do not have enough evidence to conclude that Age_group_numeric has a significant effect on StripSearch_Count after controlling for Perceived_Race. The effect size measure np2 (partial eta squared) for Perceived_Race is 0.146717, indicating that about 14.67% of the total variance in StripSearch_Count is explained by Perceived_Race after controlling for Age_group_numeric. The effect size for Age_group_numeric is smaller at 0.014487.

Source	SS	DF	F	p-unc	np2
Perceived_Race	3.340052e+05	7	5.428516	0.000009	0.146717
Age_group_numeric	2.855479e+04	1	3.248665	0.072844	0.014487
Residual	1.942524e+06	221	NaN	NaN	NaN

Table 1: Results from the ANCOVA for perceived race, age group, and strip search count

One key finding from both Table 1 and 2 is that the Perceived_Race variable has a significant effect on the StripSearch_Count after controlling for the effect of the covariate Age_group_numeric. This is evident from the significant p-values (less than 0.05) for the different levels of Perceived_Race, such as “East/Southeast Asian”, “Indigenous”, “Latino”, “Middle-Eastern”, and “South Asian” (refer to Table 2). The Age_group_numeric variable does not have a significant effect on StripSearch_Count as the p-value is 0.073 (greater than the 0.05 significance level). The R-squared value is 0.157, indicating that about 15.7% of the variance in StripSearch_Count is explained by the model, which includes both Perceived_Race and Age_group_numeric. Overall, the results of the first one-way ANCOVA suggests that there are significant differences in the strip search count based on the perceived race of individuals, even after adjusting for the effect of their age group.

Predictor	Coefficient	Std. Error	t-value	P> |t|	[0.025]	[0.975]
Intercept	107.0180	22.471	4.763	0.000	62.734	151.302
C(Perceived_Race)[T.East/Southeast Asian]	-70.4286	25.057	-2.811	0.005	-119.809	-21.048
C(Perceived_Race)[T.Indigenous]	-71.8571	25.057	-2.868	0.005	-121.238	-22.477
C(Perceived_Race)[T.Latino]	-78.2961	25.059	-3.124	0.002	-127.682	-28.910
C(Perceived_Race)[T.Middle-Eastern]	-74.4643	25.057	-2.972	0.003	-123.845	-25.084
C(Perceived_Race)[T.South Asian]	-73.2500	25.057	-2.923	0.004	-122.631	-23.869
C(Perceived_Race)[T.Unknown or Legacy]	-66.8445	24.446	-2.734	0.007	-115.021	-18.668
C(Perceived_Race)[T.White]	25.3230	24.443	1.036	0.301	-22.849	73.495
Age_group_numeric	-0.6097	0.338	-1.802	0.073	-1.276	0.057

Table 2: OLS Regression Results for the Effect of Perceived Race and Age Group on StripSearch Count

One-way ANCOVA #2

The null hypothesis for the second one-way ANCOVA is there is no significant difference in the mean StripSearch_Count between the different levels of the ‘sex’ variable (i.e., the mean StripSearch_Count is the same for different sexes) after accounting for the effect of Age_group_numeric. The alternative hypothesis is that there are significant differences in the mean for StripSearch_Count for different sexes after accounting for the effect of Age_group_numeric.

The results of the second one-way ANCOVA that was conducted can be found in Table 3. The p-value for the main effect of Sex was 0.003 (less than 0.05), therefore there is statistically significant evidence to reject the null hypothesis for the main effect. This indicates that there is a significant difference in the mean StripSearch_Count between different sexes after controlling for the effect of ‘Age_group_numeric”. The p-value for the covariate, Age_group_numeric, was

0.07, therefore since the p-value is greater than 0.05 there is not enough evidence to reject the null hypothesis. This indicates that there is no significant linear relationship between Age_group_numeric and StripSearch_Count after controlling for the effect of Sex.

Source	SS	DF	F	p-unc	np2
Sex	1.171080e+05	2	6.128126	0.002560	0.051441
Age_group_numeric	3.095154e+04	1	3.239317	0.073224	0.014131
Residual	2.159421e+06	226	NaN	NaN	NaN

Table 3: Results from the ANCOVA for sex, age group, and strip search count

Prediction Interval

In the context of our regression analysis, we computed prediction intervals to provide a range where we expect future observations to fall with a specified level of confidence. We used the OLS regression model to predict the likelihood of individuals being subjected to a strip search (StripSearch) based on their race (Non_White), gender (Sex), and age group (Age_group_numberic). The data for this analysis consisted of 64, 606 observations with complete information on the variables of interest. The dataset was split into training and test set, with 80% of the data used for model training and 20% for model evaluation.

The results of the OLS Regression that was conducted can be found in Table 4. The coefficient for Non_White is -0.0203, suggesting that being non-white is associated with a decrease in the likelihood of being subjected to a strip search compared to being white. The coefficient for Sex is 0.0244, indicating that, holding other variables constant, males (represented by Sex = 1) are more likely to be subjected to a strip search compared to females (represented by Sex = 0), with an increase in the predicted likelihood of approximately 2.44 percentage points. The coefficient for Age_group_numeric is -0.0014, suggesting that as individuals' ages increase by one unit (e.g., from 29.5 to 30.5), the predicted likelihood of being subjected to a strip search decreases by approximately 0.14 percentage points, while keeping other variables constant. The p-value for Non-white, Sex, age group are all 0.000, indicating that this predictor is statistically significant at common significance levels.

Predictor	Coefficient	Std. Error	t-value	P>|t|	[0.025	0.975]
Intercept (const)	0.1563	0.005	28.675	0.000	0.146	0.167
Non_White	-0.0203	0.003	-7.088	0.000	-0.026	-0.015
Sex	0.0244	0.004	6.883	0.000	0.017	0.031
Age_group_numeric	-0.0014	0.000	-12.524	0.000	-0.002	-0.001

Table 4: Summary of OLS Regression Results for Predicting Strip Search

We computed prediction intervals for the test data to assess the uncertainty in our model's predictions. Specifically, we calculated the mean predicted value, the 95% confidence interval (CI) around the mean, and the prediction interval (PI) for each observation in the test set.

We found that 88.66% of the prediction intervals contain the true target (the actual likelihood of strip search). This means that our model's predictions have a fairly high coverage rate, capturing the majority of true outcomes within the computed prediction intervals. This high coverage rate gives us confidence in using the model for prediction purposes.

Logistic Regression

Assumption Check

To run a logistic regression the observations must be independent of each other. Based on how the data was collected by the TPS we assume that there is no relationship between the observations within or between the groups. Additionally, there should be an adequate number of observations for each independent variable in the dataset. Sticking to the rule of thumb, we determined that we had a large sample size since the total number of observations in the dataset is greater than 500.

In order to check for the absence of multicollineraity the variance inflation factor (VIF) was calculated (Table 5). The smallest possible value for VIF is 1, suggesting a complete absence of collinerity. There are no VIF values that exceed 5 or 10 therefore the assumption is fulfilled.

VIF Factor	Features
1.041769	Non_white
1.007515	Sex_M
1.248662	Age_group_17 and under
1.912852	Age_group_18 - 24
2.704545	Age_group_25 - 34
2.265081	Age_group_35 - 44
1.701303	Age_group_45 - 54
1.367963	Age_group_55 - 64
1.105193	Age_group_65 and older

Table 5

Logistic Regression Analysis

In the logistic regression analysis, we built a model to understand the relationship between one or more independent variables (which could be categorical or continuous) and a binary dependent variable. The logistic regression model estimates the log-odds of the probability of the dependent variable being in one of the two binary classes, given the values of the independent variable(s). The purpose of the logistic regression for this research paper is to analyze the relationship between the probability of a binary outcome (StripSearch) and one or more predictor variables (Non_White, Sex, and Age_group_numeric).

According to Table 6, the slope for all but one (Age 45 - 54) of the variables of interest are significant. The coefficient for the 'Age 45 - 54' variable is not statistically significant (p = 0.263) suggesting that there is not enough evidence to conclude that being in this age group has a significant effect on being strip searched. The P-value for the slopes that are significant are much less than 0.05. Based on the previous analyses conducted the results are not surprising for the relationship between race and sex. However, the relationship between age and strip search is surprising as the previous analyses showed no evidence of there being a relationship. However, this may be due to the fact that for the ANCOVA we did not separate the ages into categories. The slope for the predictor variables are all positive except for 'Age 45 - 54' and 'Age 65 and older'. Specifically, the coefficient for the 'Age 65 and older' variable is negative and statistically significant which suggests that being in this age group is associated with a significant decrease in the likelihood of being strip searched. The positive slope means that as the predictor variable increases the probability of the event, getting strip searched, also increases.

Variable	Coef	Std Err	Z	P>|z|	[0.025	0.975]
Intercept	-2.4287	0.081	-30.147	0.000	-2.587	-2.271
Sex_M	0.2314	0.037	6.213	0.000	0.158	0.304
Non_white	-0.2027	0.029	-7.081	0.000	-0.259	-0.147
Age_group_18 - 24	0.4442	0.081	5.502	0.000	0.286	0.603
Age_group_25 - 34	0.4065	0.077	5.282	0.000	0.256	0.557
Age_group_35 - 44	0.3774	0.078	4.828	0.000	0.224	0.531
Age_group_45 - 54	0.0936	0.084	1.119	0.263	-0.070	0.258
Age_group_55 - 64	-0.2108	0.097	-2.175	0.030	-0.401	-0.021
Age_group_65 and older	-1.2888	0.205	-6.280	0.000	-1.691	-0.887

Table 6: Results from the logistic regression

Variable	Lower CI	Upper CI	OR
Intercept	0.075279	0.103233	0.088155
Sex_M	1.171656	1.355842	1.260389
Non_white	0.772023	0.863673	0.816563
Age_group_18 - 24	1.331077	1.826681	1.559312
Age_group_25 - 34	1.291297	1.745987	1.501528
Age_group_35 - 44	1.251280	1.699873	1.458429
Age_group_45 - 54	0.932096	1.293792	1.098153
Age_group_55 - 64	0.669736	0.979392	0.809897
Age_group_65 and older	0.184327	0.412055	0.275595

Table 7: The odds ratio and 95% confidence interval based on the logistic regression

Model Performance

The model's performance was assessed using metrics such as accuracy, confusion matrix, precision, recall, and F1-score. Additionally, the impact of adjusting the classification threshold on model performance was explored through a series of plots. The chosen threshold value plays a crucial role in determining the classification boundary between the positive and negative classes.

To identify an appropriate threshold, the model's accuracy and F1-score were plotted against a range of threshold values (Figure 9), allowing for the examination of how these performance metrics vary as the threshold changes.While the accuracy remains relatively stable across thresholds, the F1-score varies considerably. The F1-score balances precision and recall, making it a useful metric to assess the model's performance when the positive class is of interest.

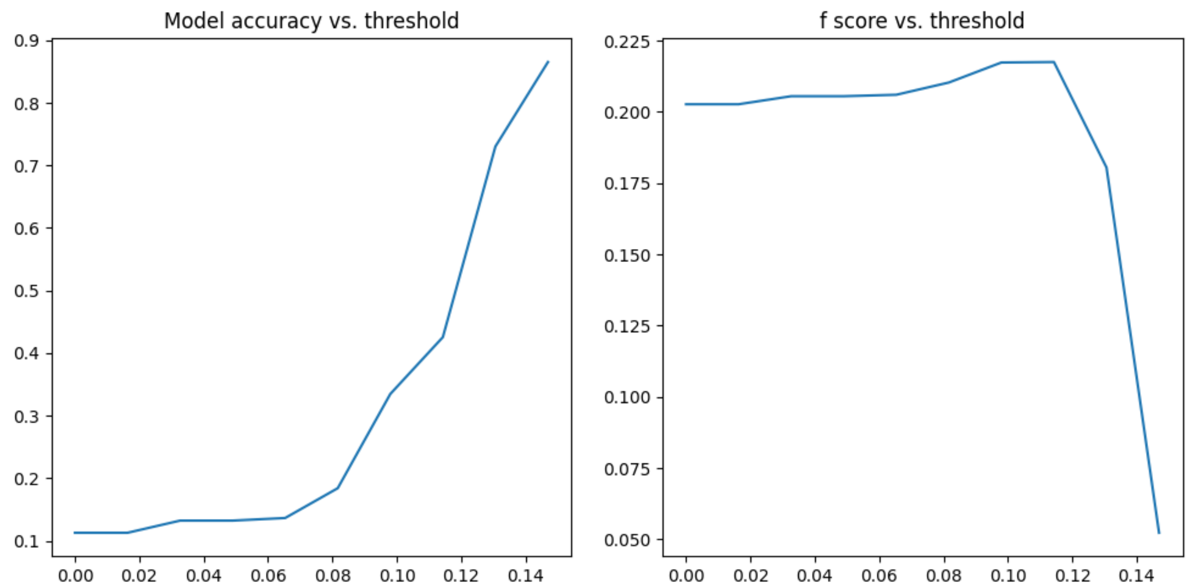

Figure 9: Accuracy and F1-Score by Classification Threshold

Based on the analysis, a threshold of 0.147 was selected to optimize performance. Using this threshold, the model achieved an accuracy of 86.53% on the training set and 86.23% on the testing set. Although the accuracy was relatively high, the number of true positive instances (TP) remained low, suggesting room for improvement in detecting positive instances. The confusion matrix for the testing set, for example, showed 52 true positives (TP) compared to 1,450 false negatives (FN). In summary, the model demonstrates satisfactory accuracy.
The confusion matrix is as Figure 10.

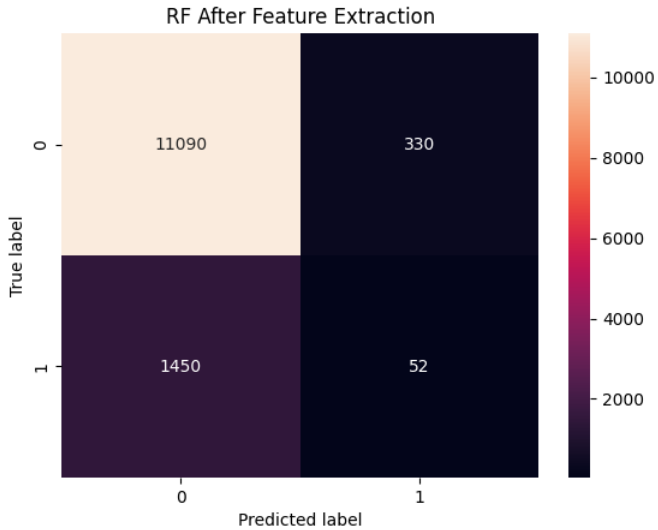

Figure 10: confusion matrix

The plot (Figure 11) shows a logistic regression curve with 95% prediction intervals. The curve represents the predicted probability of a certain outcome based on the variable "Sex_M," which likely represents a binary variable for gender. In this case, we set the 'Non_white' variable to 1, which means we're examining the case where the individual is white. For age groups, we set the variable (Age_group_25 - 34) to 1 and all other age group variables to 0, which means we're looking at the relationship for non white individuals who are in the 25-34 age group.Based on the plot, we can observe the following: For lower values of "Sex_M" (around -3), the predicted probability of the outcome is close to 0. As "Sex_M" increases, the predicted probability increases, and the curve plateaus near a predicted probability of 1 for higher values of "Sex_M" (around 3). The 95% prediction intervals are relatively narrow, suggesting a reasonable level of confidence in the predictions.

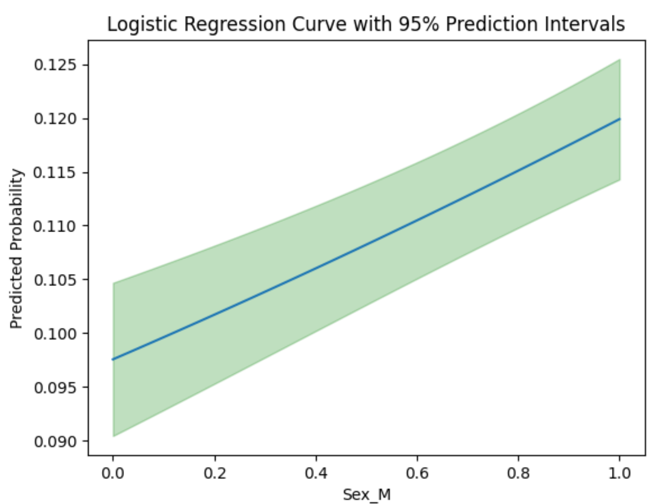

Figure 11: Logistic Regression Curve

Discussion

The analyses carried out in the results section helped to provide insights into factors associated with strip search rates, the variability in strip search rates across different demographic groups, and the accuracy of predictions for future strip search occurrences. To test this hypothesis, the study employed various statistical measures, including Cohen's D and ANCOVA.

Male vs. Female

The results from the analyses conducted suggest that there is a significant difference in strip search rates between males and females, these results can be found in the logistic regression, ANCOVA, and prediction interval. However, the effect size suggests that there is a medium effect, therefore the practical significance of this difference may be somewhat limited. According to the odds ratio, holding all the other variables constant, the odds of getting strip searched increased by 26.04% for males compared to females.

It is not surprising that the results of our analysis demonstrated that there is a difference between men and women getting strip searched. Strip searches are typically conducted based on an individual's perceived risk factor. Men are often considered a higher risk due to a higher rate of arrests, convictions, and repeat offenses. It is also thought that men are more likely to conceal weapons and contraband due to their greater physical strength and a higher probability of escape. However, this practice can be seen as discriminatory. It should be noted that there are a larger number of males included in the dataset in comparison to females and unknown, therefore this could have had an effect on the results of the one-way ANOVA.

The Toronto Police Service (TPS) states that all searches, including strip searches, must be conducted in accordance with the law, with respect for the individual's privacy and dignity, and in accordance with the guidelines set out in the Ontario Police Services Act (2019). The TPS acknowledges that males may be subject to more extensive searches due to the higher rate of crime involving males (2019). However, the TPS has taken steps to ensure that the searches are conducted in a fair and even-handed manner, regardless of gender (2019). The TPS has implemented a policy that requires all searches to be authorized by a supervisor before being performed, and that all searches must be conducted in accordance with the law (2019).

White vs. Non-White

The results from the research paper suggest that there is a significant difference in strip search rates between white and non-white individuals. However, the effect size is small, which suggests that the practical significance of this difference may be limited. The results from the ANCOVA test indicated that there was a significant difference in the mean strip search count between the different levels of race, even after controlling for age group.

There is an enormous amount of articles available that focus on the disparity in strip search rates based on race. Research has even demonstrated that people of color, such as Black and Middle Eastern individuals, are more likely to be stopped and searched by police than White individuals due to the use of racial profiling. The disparity in strip search rates between minority races and Whites can be based on social and historical context in which policing takes place. For example, there may be disparities in policing resources and practices in different neighborhoods which can lead to a disproportionate impact on minority communities. The reasons for the disproportinality between White and non-White individuals are complex and multifaceted, and may be influenced by factors such as racial profiling, implicit bias, and systemic racism within law enforcement agencies.

Age

The research paper aimed to investigate the relationship between age and strip search rates while controlling for the effects of other demographic variables. The results of the logistic regression demonstrated that the youngest age group, 18 to 24 years, had the highest mean strip search count compared to the other age groups. This is not surprising due to the high frequency of media reports about younger people being strip searched by the police, specifically black youth are most often strip searched (Burke, 2023; Dodd, 2023). The oldest age group, 65 years and above, had the lowest mean strip search count compared to the other age groups. This is most likely because older individuals are perceived as less likely to engage in criminal activities, and therefore, are less likely to be subjected to strip searches. Older people are also often perceived as less of a threat to law enforcement compared to younger people, such as youth.

The relationship between age and strip search from the logistic regression is surprising based on the specific context of the data, the research question, and the findings from both the ANCOVA and logistic regression analyses. It's possible that in the ANCOVA analysis, there was no significant evidence of a relationship between age and the dependent variable of interest (e.g., strip search) after controlling for other factors. However, the logistic regression analysis may have found a significant relationship between age and strip search. This discrepancy could be surprising because it may indicate that the relationship between age and strip search is more complex than initially thought and may depend on other factors or interactions that were captured in the logistic regression model but not in the ANCOVA analysis.

Conclusion

The findings help to contribute to a better understanding of the patterns and potential disparities in strip searches. The findings from this study have important implications for the Toronto government as they suggest that there may be disparities in how strip searches are conducted across different demographic groups. Addressing these disparities is crucial for ensuring that the Toronto Police operate in a fair and unbiased manner. This research paper provides valuable insights into the relationship between an individual's demographic characteristics and the likelihood of being subjected to a strip search, while accounting for temporal changes. The findings suggest that race and sex are significant predictors of being subjected to a strip search, after controlling for age group and temporal changes in the strip search rates. Addressing these disparities is crucial for ensuring that law enforcement agencies operate in a fair and unbiased manner.

Future research should further look into and understand the underlying factors that may contribute to differences in strip search rates based on race, and to develop evidence based-policies and procedures to ensure that strip searches are conducted failty and without bias. This may include providing training and education to police officers, engaging with community members to build trust, and implementing reforms to reduce the use of racial profiling and other discriminatory practices.

References

Babbar, M. (2021). *Scholarly Commons: Northwestern pritzker school of law*. Site. Retrieved February 25, 2023, from https://scholarlycommons.law.northwestern.edu/

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates.

Psutka, M., & Sheehy, E. (2016, December 1). *Strip-searching of women in Canada: Wrongs and rights*. The Canadian Bar Review. Retrieved February 25, 2023, from https://cbr.cba.org/index.php/cbr/article/view/4376

Smith, B. W., Stinson, P. M., & Paternoster, R. (2020). *Racial and ethnic disparities in the American criminal justice system: A systematic review and meta-analysis. Journal of Criminal Justice, 68, 101657*. Retrieved February 25, 2023, from https://doi.org/10.1016/j.jcrimjus.2020.101657

Alexander, M. (2012). *The new Jim Crow: Mass incarceration in the age of colorblindness. The New Press*. Retrieved February 25, 2023, from https://www.thenewpress.com/books/new-jim-crow

Lemke, M. (2022, July 18). *Policing toronto: Strip searching in a divided city - the bullet*. Socialist Project. Retrieved February 25, 2023, from https://socialistproject.ca/2022/07/policing-toronto-strip-searching-in-a-divided-city/

Preventing and Addressing Sexual Abuse in Juvenile Facilities: A Call to Action. Report of the National Prison Rape Elimination Commission. (2009, June). Retrieved February 25, 2023, from https://www.ojp.gov/pdffiles1/226680.pdf

Bevans, R. (2022, December 19). *An introduction to T tests: Definitions, formula and examples*. Scribbr. Retrieved February 28, 2023, from https://www.scribbr.com/statistics/t-test/

Statistical Power Analysis. Statistics Solutions. (2021, August 10). Retrieved April 16, 2023, from https://www.statisticssolutions.com/dissertation-resources/sample-size-calculation-and-sample-size-justification/statistical-power-analysis/

What is power analysis? TIBCO Software. (n.d.). Retrieved April 16, 2023, from https://www.tibco.com/reference-center/what-is-power-analysis

One-way Ancova in SPSS statistics. How to perform a one-way ANCOVA in SPSS Statistics | Laerd Statistics. (n.d.). Retrieved April 16, 2023, from https://statistics.laerd.com/spss-tutorials/ancova-using-spss-statistics.php

One-way ancova. StatsTest.com. (2020, November 3). Retrieved April 16, 2023, from https://www.statstest.com/one-way-ancova/

Bedre, R. (2022, January 23). *Ancova using R and python (with examples and code)*. Data science blog. Retrieved April 16, 2023, from https://www.reneshbedre.com/blog/ancova.html

Prediction Interval | *Overview, Formula & Examples.* (2022, March 24). Study.com. Retrieved April 16, 2023, from https://study.com/learn/lesson/prediction-interval-overview-formula-examples.html

Law document english view. Ontario.ca. (2018, November 19). Retrieved February 28, 2023, from https://www.ontario.ca/laws/statute/s19001

Lawton, G., Burns, E., & Rosencrance, L. (2022, January 20). *What is logistic regression? - definition from Searchbusinessanalytics*. Business Analytics. Retrieved April 16, 2023, from https://www.techtarget.com/searchbusinessanalytics/definition/logistic-regression

Thanda, A., Anamika Thanda Anamika ThandaOriginally from India, Thanda, A. T. A., Thanda, A., & India, O. from. (2022, December 19). *What is logistic regression? A beginner's guide [2023]*. CareerFoundry. Retrieved April 16, 2023, from https://careerfoundry.com/en/blog/data-analytics/what-is-logistic-regression/

The Guardian. (2023, March 26). *Race disparity in police strip searches of children in England and Wales.* Retrieved from https://www.theguardian.com/uk-news/2023/mar/26/race-disparity-police-strip-searches-of-children-england-and-wales

The Mirror. (2023, April 16). *Police bullies carried out 2,800 victimising searches on children.* Retrieved from https://www.mirror.co.uk/news/politics/police-bullies-carried-out-2800-29545032

Lani, J. (2021, June 23). *Binary logistic regression*. Statistics Solutions. Retrieved April 16, 2023, from https://www.statisticssolutions.com/binary-logistic-regression/