Data Analysis of Likelihood of Strip Search and Number of Items Found at Arrest

by

Junwei Shen
Peiwen Zhang

A midterm report submitted in conformity with the
requirements for the course of INF2178H

Faculty of
Information
University of
Toronto

Table of Contents

Introduction/background

According to the guidelines of the Toronto Police Service (TPS), an arrest can be performed under the following circumstances. Firstly, if a police officer has a reasonable belief that the person has committed or is about to commit a criminal offense. Secondly, if the individual is found to be committing, or there is a reasonable belief that they are about to commit, a breach of peace. Thirdly, if the officer has a belief on reasonable grounds that the individual is wanted on a warrant of arrest or committal that is valid within the territorial jurisdiction where the person is located. A strip search is defined as a search that includes the removal of part of the clothing and a visual inspection of the body performed by the police officer based on the description of the dataset in the TPSe Public Safety data portal (2021). It is done when an accused person is being "booked," or brought into the station and is often justified by the police as necessary to recover hidden weapons or drugs and to protect the accused, other inmates, or officers.

The TPS has specific guidelines and protocols for conducting strip searches, including requirements for the presence of at least two officers of the same gender as the person being searched and the use of appropriate search equipment. Regardless of those restrictions, the TPS has been a topic of concern and facing intense scrutiny over the years for their use of strip searches. According to Gills (2019), "allegations of racial profiling, excessive force, and violations of human rights have led to calls for reform and greater transparency in police practices." The TPS modified the code of conduct on strip searches on November 4th, 2021. Data had shown that around 27% of arrests resulted in a strip search before the policy change, while only around 4.9% of arrests resulted in a strip search after changing the policy (Gills, 2019).

In this report, our group will use the data from the TPS Public Safety data portal to analyze the patterns and trends of the likelihood of strip searches, the number of strip searches, and the number of items found at arrest. Various other relevant sources, including media articles, and government guidelines were referenced to provide a background on this topic. This report aims to provide an overview of the Toronto police arrests, strip searches, strip search likelihood, and items found analysis, highlighting some key findings, contributing to ongoing discussions, and practice for data analysis.

1.1 Research Questions

Research Question 1: What is the association between outcome items found and explanatory variable perceived race after controlling for the covariate, the number of strip searches?
Research Question 2: What is the association between outcome items found and explanatory variable sex after controlling for the covariate, the number of strip searches?
Research Question 3: What is the association between the binary outcome likelihood of being strip searched and explanatory variables perceived race, sex, age group, and the number of actions at arrest?

Literature Review

The study will make use of three prior studies pertaining to the matter of Toronto arrests and strip searches, in order to enhance the comprehension of the issue at hand and to clarify the objective of the research.

The first study is *A collection Impact: Interim report on the inquiry into racial profiling and racial discrimination of Black persons by the Toronto Police Service* (2018) by the institution of Ontario Human Rights Commission (OHRC), which involved a methodical analysis and examination of research conducted on the apprehension of black individuals during the years 2013 to 2017. The study conducted a qualitative analysis and found that a Black person in Toronto was nearly 20 times more likely than a White person to be involved in a fatal shooting by the TPS. In addition, in the data collected by the OHRC from the Special Investigations Unit (SIU), it was found that Black individuals were disproportionately represented in cases involving the use of force (28.8%), shootings (36%), deadly encounters (61.5%), and fatal shootings (70%). Despite only constituting 4.1% of Toronto's population, Black men were complainants in 25% of SIU cases involving sexual assault allegations against TPS officers. Overall, the report arrived at the conclusion that the TPS displays racial prejudice, and prompt action is imperative to confront systemic racism within the TPS and ensure that everyone is treated justly and equitably.

The second study *"Clearly, we were doing it wrong": Toronto police are doing far fewer strip searches under strict new rules, Interim chief says* (2020) by Wendy G. investigated the outcomes of strip searches after the TPS implemented new rules and guidelines. The article highlighted that the implementation of the new policy has resulted in a substantial reduction in

the frequency of strip searches performed by the TPS and has fostered better relations with the community. Additionally, the article emphasized the importance of ongoing monitoring and supervision to ensure that the TPS follows the new policy and does not engage in any discriminatory conduct.

The third study is a video created by Toronto Police Service RBDC (Race and Identity-Based Data Collection), the video is titled *Strip Searches: Measurement & Outcomes*. The study consists of 31,979 arrests data in 2020, and 17,000 were booked into custody at the police station with 7,114 strip searches. Note that strip searches account for 22.2% of all arrests. The study stated that "items were found in 40% of strip searches conducted" (Toronto Police Service, 2022). It highlights that strip search rates and items found rates were related to the type of offenses, of which drug-related offenses accounts for the highest rates. Another interesting result provided by the study is that although Black people are more likely to be strip searched at arrest than White people (48% vs. 38%), Black people have fewer proportions of items found compared to White people (38% vs. 50%). This reveals an insight that perceived race might have influences on strip search rate and items found rate, and these actions potentially cause unnecessary searches.

Dataset Description

2.1 Dataset

This report will reference the arrests and strip searches dataset, titled "Arrests_and_Strip_Searches_(RBDC-ARR-TBL-001)", from the TPS Public Safety data portal with around 65276 records containing the date range from the beginning of 2020 to the November of 2021. The dataset contains age at arrest, sex, perceived race, demographic information on the individual arrested, types of charges they faced, and the circumstances surrounding their arrest. Besides that, the dataset contained whether the arrested individuals were being stripped searched, the reasons for strip searching, and whether items were found from the strip search.

This dataset provided transparency and accountability regarding TPS's use of these practices and allowed for the analysis of trends and patterns over time. It is an essential resource for analyzing the arrests trend, strip search trend, and the rate trend between strip search and arrests. Click here to view the original dataset.

2.2 Data preparation

During the data preparation process, our group formatted two datasets. We first created a cleaned dataset and followed the procedure of removing duplicates, dealing with missing values, and standardizing and reformatting the dataset. We found 4 duplicated entries based on the distinct values of *eventID*, *arrestID*, and *personID*. For the Search Reason columns, we filled in missing values as 0. Additionally, we dropped entries that were null in the *Age_group__at_arrest_* column and *Perceived_Race* column since the percentage of missing values was relatively small.

Our first dataset was created directly by copying the cleaned dataset with summing *Actions_at_arrest___Concealed_i, Actions_at_arrest___Combative__*, *Actions_at_arrest___Mental_inst*, *Actions_at_arrest___Concealed_i*, *Actions_at_arrest___Assaulted_o,* and *Actions_at_arrest___Cooperative* as the number of *Actions_at_arrest.* This dataset is for the usage of logistic regression analysis.

The second dataset was reformatted the dataset by summing the number of items found as the total number of items found and summing the number of strip searches as the total number of strip searches performed. We grouped by *Arrest_Year, Arrest_Month*, *Sex*, and *Perceived_Race* columns to perform these calculations. This dataset is for the analysis of ANCOVA.

2.3 Measurement

2.3.1 Independent Variables

After identifying the research problems, we have defined different independent variables for different research questions.

For research question 1&2, we have defined the independent variables of this research as the month when the sex identified by the individual (*Sex*), the race recorded by the individual (*Perceived_Race*), and the number of strip searches when the individual was arrested (*StripSearch*). Both *Sex* and *Perceived_Race* are categorical variables with *Sex* having two categories, and *Perceived_Race* having eight categories.

For research question 3, we have specified the independent variable to be the sex identified by the individual (*Sex)*, the race recorded by the individual (*Perceived_Race*), the age group of the individual (*Age*), and the number of actions when the individual was arrested(*Actions_at_arrest*).

2.3.2 Dependent Variables

For research questions 1&2, the dependent variable is the number of items found from individuals after being strip searched (*ItemsFound*). To be specific, items found is a binary outcome (whether items are found/not found on an individual), so as we sum all the items found, the depend variable actually means the occurrence of the event "items found" (i.e. actually finding items on the individual, encoded as "1"). However, for simplicity and clarity, we decided to refer to this term as "number of items found" for the following sections and discussions.

The dependent variables for this project are whether the individual was subjected to a strip search or not (*StripSearch*).

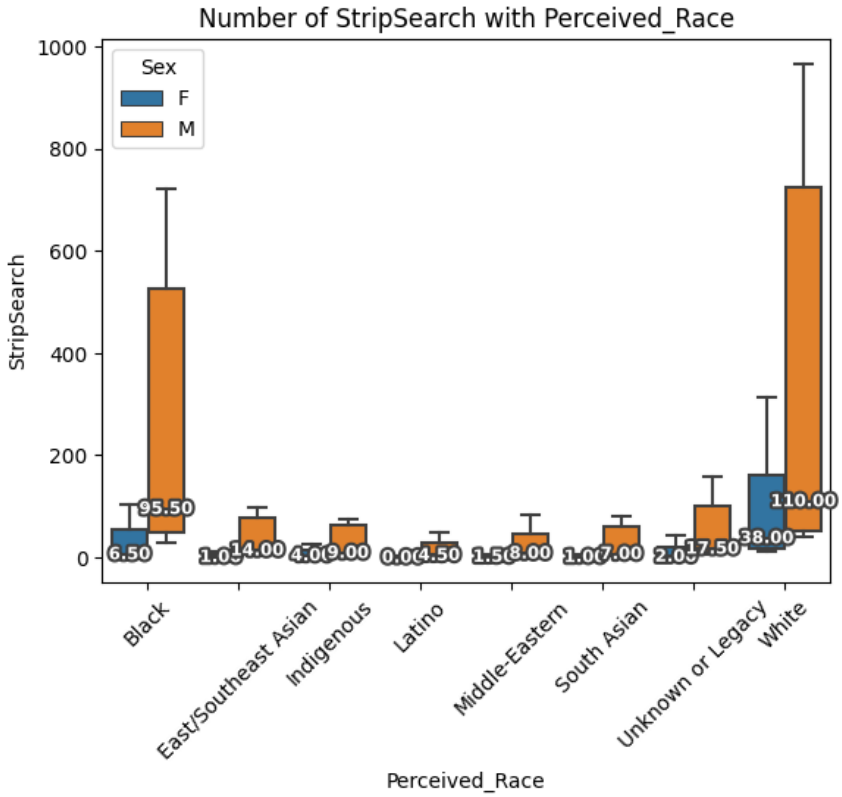
Descriptive Analysis

3.1 Research Questions 1&2 EDA

We established this EDA to investigate Research Questions 1 & 2. Box plots and Interaction plots are utilized to study the relationships between outcomes mean strip search number and items found with respect to explanatory variable perceived race. The sex of individuals is also included in the plots to reveal interactions between perceived race and sex.

3.1.1 Box Plots

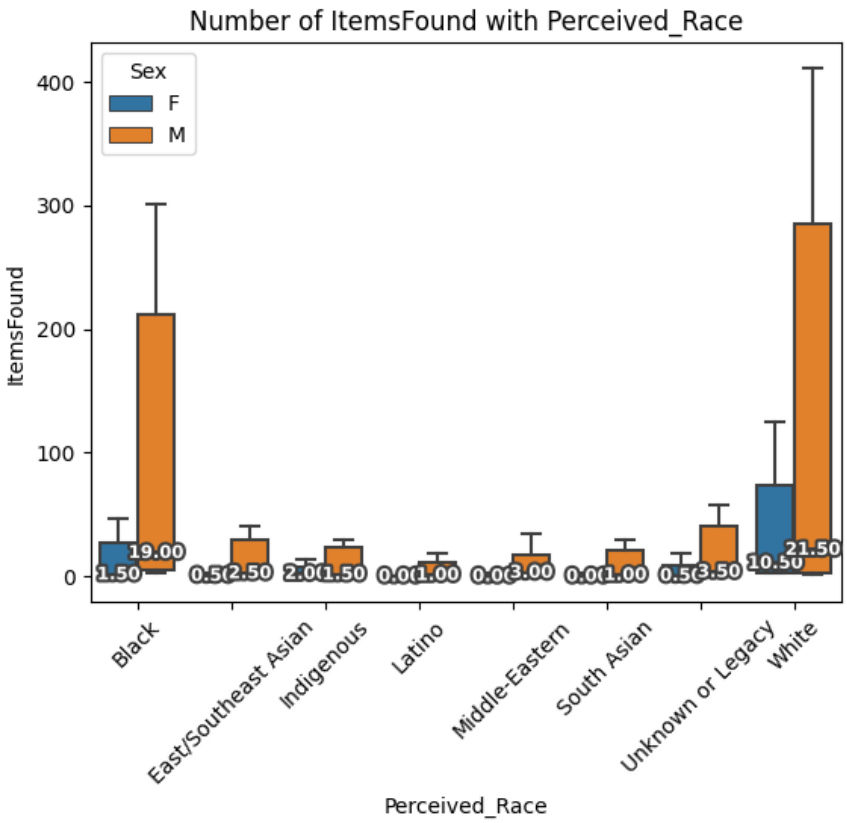
Figure 3.1.1: Box plot of Mean Strip Search Number vs. Perceived Race and Sex



In Figure 3.1.1, the box plot reveals the relationship between mean strip search numbers and perceived race & sex. The numbers in the middle of the box and whisker plot represent the median of the mean strip search number within the specific race. It is worth noting that: (a) all "Male" individuals have a higher median of mean strip search numbers than "Female" individuals in all races; (b) Individuals who are labeled as "White Male" has the greatest maximum mean strip search number; (c) For "Male" Individuals who are labeled as "White" and

"Black", the medians of mean strip search number are 10 times greater than individuals who are labeled as "Indigenous".

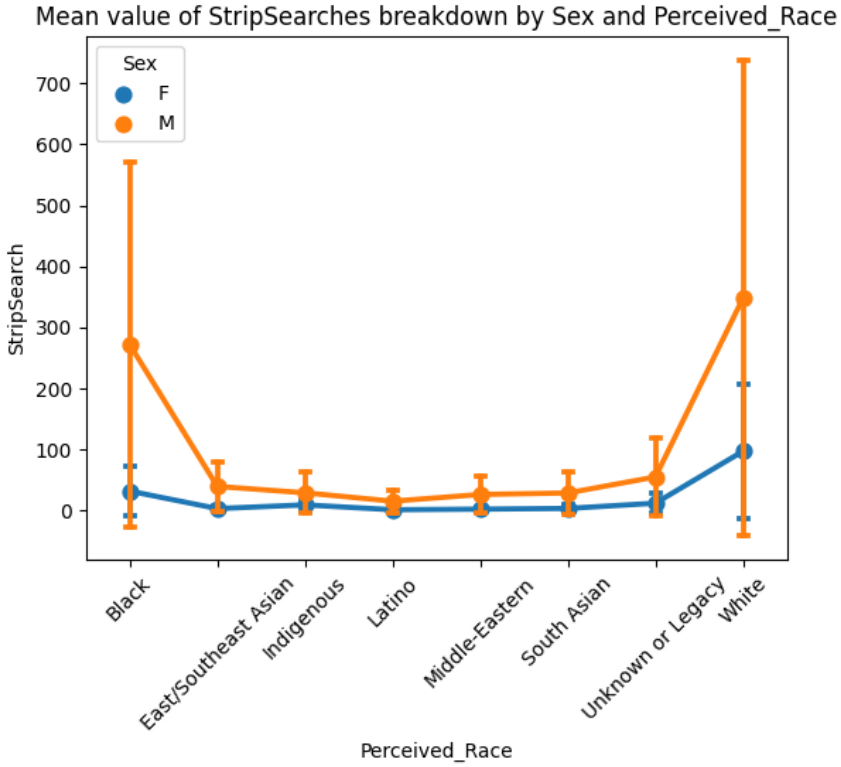
Figure 3.1.2: Box plot of Mean Items Found vs. Perceived Race and Sex



In Figure 3.1.2, the box plot reveals the relationship between the average number of items found and perceived race & sex. The numbers in the middle of the box and whisker plot represent the median of the mean strip search number within the specific race. It is worth noting that: (a) all "Male" individuals have a higher median of the average number of items found than "Female" individuals in all races; (b) Individuals who are labeled as "White Male" has the greatest maximum average number of items found; (c) For "Male" Individuals who are labeled as "White" and "Black", the medians of the average number of items found are 20 times greater than individuals who are labeled as "Indigenous" and "Latino".

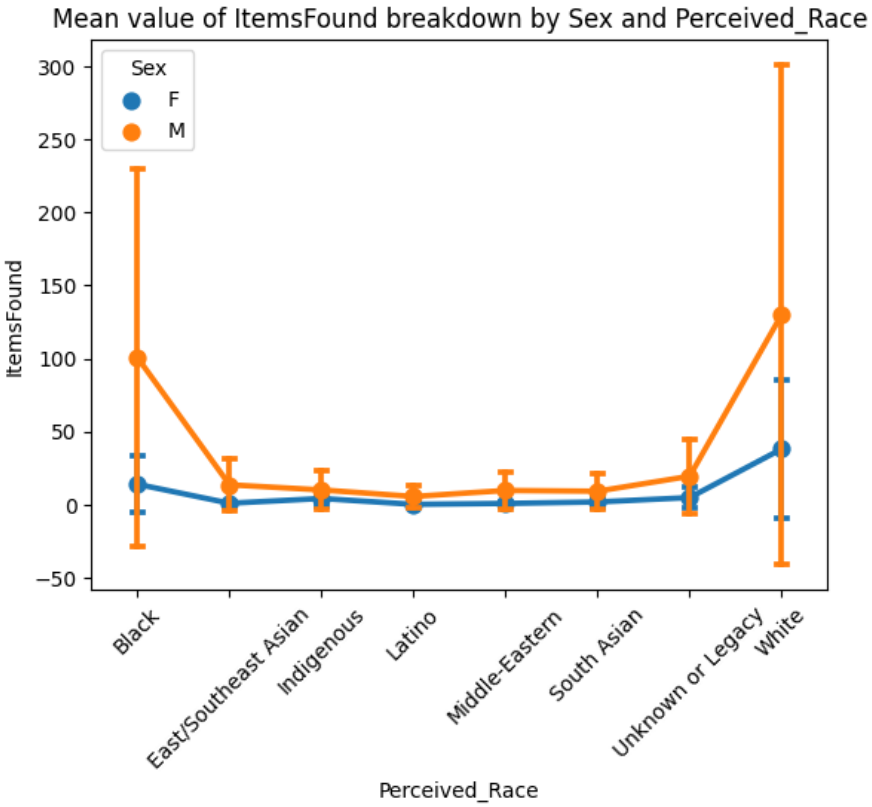
3.1.2 Interaction Plots

Figure 3.1.3: Interaction plot of Mean Strip Search Number vs. Perceived Race and Sex



In Figure 3.1.3, while the interaction plot does not provide any information on the statistically significant difference, the plot showed: (a) “Male” individuals have a higher mean strip search number compared to “Female”; (b) there is a noticeable big difference in mean strip search numbers between “Male” and “Female” individuals who are identified as “White” and “Black”. (c) “Male” individuals who are identified as “White” and “Black” have much higher mean strip search numbers compared to other races in both “Male” and “Female”

Figure 3.1.4: Interaction plot of Mean Items Found vs. Perceived Race and Sex



In Figure 3.1.4, while the interaction plot does not provide any information on the statistically significant difference, the plot showed: (a) “Male” individuals have a higher average number of items found compared to “Female”; (b) there is a noticeable big difference in the average number of items found between “Male” and “Female” individuals who are identified as “White” and “Black”. (c) “Male” individuals who are identified as “White” and “Black” have a much higher average number of items found compared to other races in both “Male” and “Female”

3.2 Research Question 3 EDA

To begin our discussion on Research Question 3, we established this EDA in preparation to the logistic regression and associated hypothesis testing. Pivot tables, bar plots, and pie charts are constructed to investigate the relationship among mean strip search rates concerning race, sex, age, and the number of actions at arrest. The interactions between different independent variables are also considered for counting for potential correlation.

3.2.1 Pivot Tables

In the following tables we studied the outcome mean strip search rates with respect to various explanatory variables, such as race, sex, age, and the number of actions at arrest. Notice the mean strip search rates are derived from the Number of Strip Searches grouped by categories for a specific explanatory variable divided by the total group size for that category.

Table 3.2.1: Mean Strip Search Rate vs. Perceived Race

Perceived_Race	MeanStripSearchRate	StripSearchers	GroupSize
Black	0.138902	2433	17516
East/Southeast Asian	0.077289	341	4412
Indigenous	0.158385	306	1932
Latino	0.074703	132	1767
Middle-Eastern	0.070436	228	3237
South Asian	0.071132	257	3613
Unknown or Legacy	0.106097	536	5052
White	0.128672	3565	27706
Total	0.119537	7798	65235

In Table 3.2.1, we noticed that the mean strip search rates are higher among individuals whose perceived race is “Indigenous” and “Black”, with “Indigenous” being the highest.

Table 3.2.2: Mean Strip Search Rate vs. Sex

Sex	MeanStripSearchRate	StripSearchers	GroupSize
F	0.101769	1283	12607
M	0.123793	6515	52628
Total	0.119537	7798	65235

In Table 3.2.2, we noticed that the mean strip search rate is the highest for the sex “Male” compared to “Female”.

Table 3.2.3: Mean Strip Search Rate vs. Age

Age	MeanStripSearchRate	StripSearchers	GroupSize
Aged 17 years and under	0.006613	9	1361
Aged 17 years and younger	0.161214	271	1681
Aged 18 to 24 years	0.134389	1349	10038
Aged 25 to 34 years	0.132270	2770	20942
Aged 35 to 44 years	0.129080	2096	16238
Aged 45 to 54 years	0.099835	905	9065
Aged 55 to 64 years	0.078901	362	4588
Aged 65 and older	0.049679	31	624
Aged 65 years and older	0.007163	5	698
Total	0.119537	7798	65235

In Table 3.2.3, we noticed that the mean strip search rate is the highest for the age group "Aged 17 years and younger", where the age group "Aged 18 to 24 years" and "Aged 25 to 34 years" remain higher compared to the rest of the groups.

Table 3.2.4: Mean Strip Search Rate vs. Actions at Arrest

Actions_at_arrest	MeanStripSearchRate	StripSearchers	GroupSize
0	0.111070	3566	32106
1	0.117172	3682	31424
2	0.309686	454	1466
3	0.413146	88	213
4	0.307692	8	26
Total	0.119537	7798	65235

In Table 3.2.4, we noticed that the mean strip search rate is the highest for the individuals whose number of actions at arrest is "3", meaning they have performed 3 different types of actions at arrest.

Table 3.2.5: Mean Strip Search Rate vs. Sex & Actions at Arrest

Sex	Actions_at_arrest	MeanStripSearchRate	StripSearchers	GroupSize
F	0	0.090852	576	6340
	1	0.103762	615	5927
	2	0.264516	82	310
	3	0.384615	10	26
	4	0.000000	0	4
M	0	0.116044	2990	25766
	1	0.120289	3067	25497
	2	0.321799	372	1156
	3	0.417112	78	187
	4	0.363636	8	22
Total		0.119537	7798	65235

In Table 3.2.5, we studied the interaction between sex and the number of actions at arrest, and we noticed that the mean strip search rate is the highest for both “Female” and “Male” whose number of actions at arrest is “3”, with “Male” has higher mean strip search rate than “Female” at all possible number of actions at arrest. Interestingly, “Female” does not commit 4 different types of actions at arrest.

Table 3.2.6: Mean Strip Search Rate vs. Sex & Age

Sex	Age	MeanStripSearchRate	StripSearchers	GroupSize
F	Aged 17 years and under	0.014184	4	282
	Aged 17 years and younger	0.059867	27	451
	Aged 18 to 24 years	0.103915	215	2069
	Aged 25 to 34 years	0.123433	522	4229
	Aged 35 to 44 years	0.111868	345	3084
	Aged 45 to 54 years	0.084270	135	1602
	Aged 55 to 64 years	0.045588	31	680
	Aged 65 and older	0.037037	4	108
	Aged 65 years and older	0.000000	0	102
M	Aged 17 years and under	0.004634	5	1079
	Aged 17 years and younger	0.198374	244	1230
	Aged 18 to 24 years	0.142301	1134	7969
	Aged 25 to 34 years	0.134506	2248	16713
	Aged 35 to 44 years	0.133115	1751	13154
	Aged 45 to 54 years	0.103176	770	7463
	Aged 55 to 64 years	0.084698	331	3908
	Aged 65 and older	0.052326	27	516
	Aged 65 years and older	0.008389	5	596
Total		0.119537	7798	65235

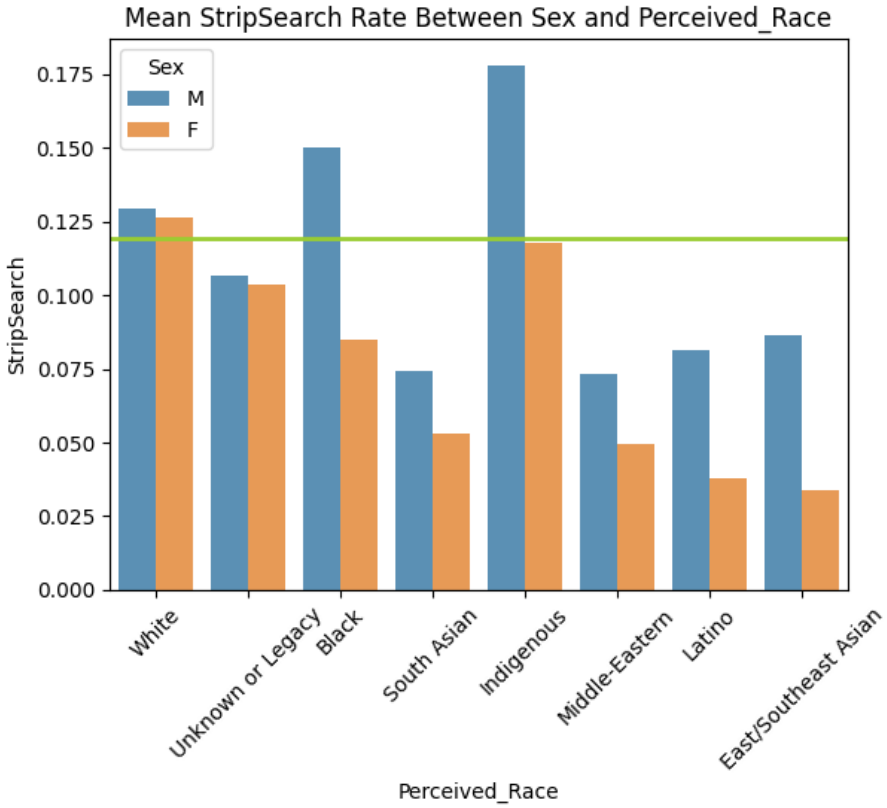
In Table 3.2.6, we studied the interaction between sex and age at arrest, and we noticed that the mean strip search rate is the highest for “Females” who are “aged 25 to 34 years” and

"Males" who are "aged 17 years and younger"."Male" has higher mean strip search rate than "Female" at all age groups. One interesting fact is that "Females" who are "aged 65 years and older" has 0 mean strip search rate.

3.2.2 Bar Plots

Bar plots were utilized to illustrate the relationship between the outcome "mean strip search rates" and explanatory variables, such as race, age, and the number of actions at arrest. In addition, the bars are colored based on the sex of the individual, in this way, interactions between race, age, and number of actions at arrest vs. sex will be revealed.

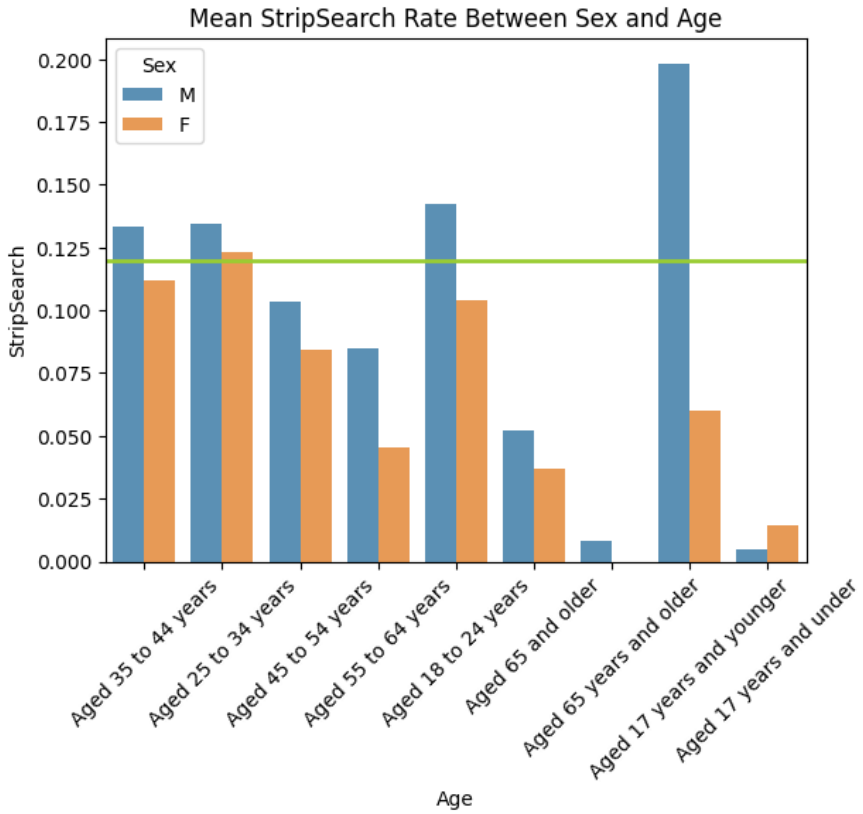
Figure 3.2.1: Bar Plot of Strip Search Rate vs. Perceived Race and Sex



In Figure 3.2.1, the bar plot illustrates the relationship between mean strip search rates and perceived race, notice that the green horizontal line represents the average mean strip search rate among all groups. It is worth noting that individuals with perceived race "Indigenous" and sex "Male" has the highest mean strip search rate among all groups, while individuals with perceived race "White" and sex "Female" has the highest mean strip search rate among all "Female" race

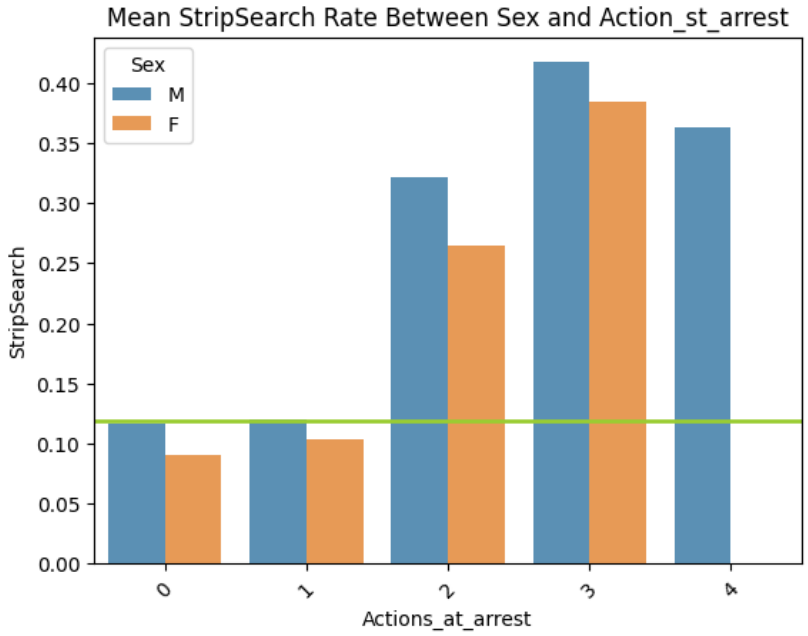
groups. The strip search rate of "Black Male", "White Male" and "White Female" also reach above the average.

Figure 3.2.2: Bar Plot of Strip Search Rate vs. Age and Sex



In Figure 3.2.2, the bar plot illustrates the relationship between strip search rates and age, notice that the green horizontal line represents the average mean strip search rate among all groups. It is worth noting that individuals from the age group "35 to 44 years", "25 to 34 years", "18 to 24 years", and "17 years and younger" with the sex "Male" has a mean strip search rate above the total average (green horizontal line), while the highest mean strip search rate appears in "Male" individual from age group "17 years and younger". Interestingly, the mean strip search rate for "Female" individuals from all age groups except "25 to 34 years" remain under the total average of the mean strip search rate.

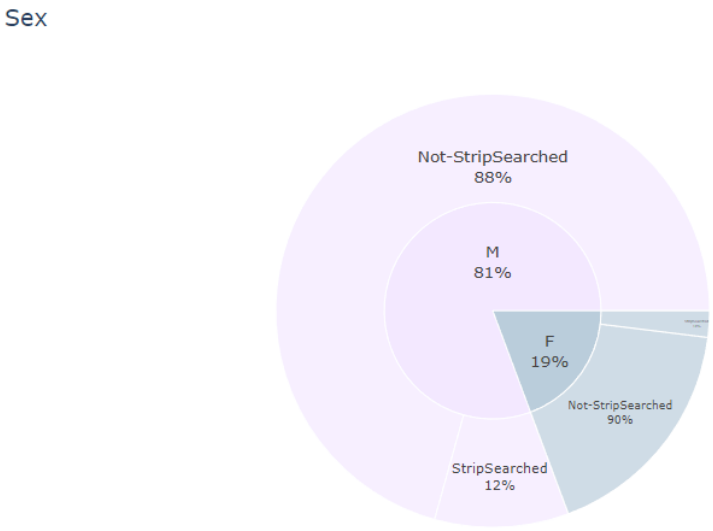
Figure 3.2.3: Bar Plot of Strip Search Rate vs. Actions at Arrest and Sex



In Figure 3.2.3, the bar plot illustrates the relationship between strip search rates and the number of actions at arrest, notice that the green horizontal line represents the average mean strip search rate among all groups. Firstly, “Male” and “Female” individuals who have committed 2,3 and 4 different types of actions at arrest all reach above the average of total mean strip search rate, while “Male” individuals who have committed 3 different types of actions at arrest have the highest mean strip search rate among all groups. Interestingly, individuals who committed 0 and 1 type of actions at arrest have similar mean strip search rates.

3.2.2 Pie Charts

Figure 3.2.4: Pie Chart of Individuals' Strip Searched/Not Strip Searched by Sex



In Figure 3.2.4, the pie chart illustrates the proportion of individuals who have been strip searched vs. individuals who have not been strip searched by different sex. Overall, this dataset contains more individuals whose sex is "Male" compared to "Female". Specifically, among all individuals, "Male" has a higher proportion of being strip searched and a lower proportion of not being strip searched compared to "Female". However, the proportion of "Male" individuals who have been strip searched does not vary much from "Female" individuals (12% vs. 10%).

Methods

In this report, the ANCOVA method, and the statistical logistic analysis method were chosen for analysis. The method was applied to different research questions.

The ANOCOVA method was employed to research questions 1 and 2 due to the reason that there is one quantitative and one categorical explanatory variable in both research questions. Additionally, the categorical variables are the treatment of primary interests while the quantitative variable is the control variable of second interest.

The statistical logistical regression method was applied to research question 3 because the dependent variable *StripSearch* is a categorical variable with the outcomes of individuals getting strip searched and not being strip searched. Conventionally code 1 was used to represent the individual being strip searched while code 0 was used to represent the individual not being strip searched.

Result and Findings

5.1 ANCOVA Analysis and Findings

In this section of the report, we intend to present results and findings regarding our research questions 1 and 2. Specifically, we studied the outcome number of items found with respect to explanatory variables sex and perceived race while controlling for the number of strip searches. To do so, we first established power analysis in the context of ANCOVA (Analysis of Covariance), we believe that it can help determine the sample size required to detect significant differences between groups or conditions while controlling for the effects of covariates. Moreover, we also performed ANCOVA analysis on both of the research questions and present our results on hypothesis tests and findings of the significance level.

5.1.1 Power Analysis

Sex

Table 5.1.1: Power Anlysis Results for Items Found vs. Sex

Effect size (Cohen's D)	Sample Size of items found for male	Actual Size of items found for male	Sample Size of items found for female	Actual Size of items found for female
0.4658182086199752	73.317	64	73.317	64

Prior to computing a t-test to analyze whether the number of items found (outcome variable) for an individual differed between different sex (two-level explanatory variable), we calculated the effect size of the explanatory variable using Cohen's D metric, which was 0.46.

After obtaining the effect size, the required sample size was computed using the obtained effect size and establishing the statistical power at 80%. The results indicated that a sample size of 73.317 was required for males in terms of the outcome "items found", while a sample size of 73.317 was required for females. This is significant because the sample size provided in the dataset is 64 and 64 for males and females respectively, which impacts the reliability of the results.

Perceived Race

Table 5.1.2: Power Anlysis Results for Items Found vs. Perceived Race

Effect size	Sample Size of items found for all race groups	Actual Size of items found for all race groups
0.1061836855317018	17.043322490720534	16

Prior to computing a t-test to analyze whether the number of items found (outcome variable) for an individual differed between different perceived race (eight-level explanatory variable), we calculated the effect size of the explanatory variable, which was 0.11.

After obtaining the effect size, the required sample size was computed using the obtained effect size and establishing the statistical power at 80%. The results indicated that a sample size of 17.04 was required for all race groups in terms of the outcome "items found". This is

significant because the sample size provided in the dataset is 16 for all race groups, which impacts the reliability of the results.

5.1.2 ANCOVA

Table 5.1.3: ANCOVA analysis table between Items Found and (Sex & Strip Searches)

Source	SS	DF	F	p-unc	np2
Sex	631.646501	1	19.352167	2.301684e-05	0.134062
Strip Search	490078.600071	1	15014.858759	4.504537e-132	0.991744
Residual	4079.946804	125	N/A	N/A	N/A

To begin the discussion of the first ANCOVA analysis, we first note the null hypothesis: individuals from different sex have the same numbers of items found on average after accounting for their number of strip searches.

Statistically speaking, the p-unc stands for “uncorrected p-value”, which is less than 0.05. Thus, we can reject the null hypothesis that different sex results in the same average number of items found, even after controlling for the number of strip searches. Meaning that different sex does have an impact on the average number of items found, after accounting for the number of strip searches.

Practical speaking, we hypothesized that the sex of individuals would be able to predict the average number of items found. From our results, we see that there is a statistically significant relationship between sex and the number of items found when controlling for the number of strip searches. This means that if we randomly select “Male” and “Female” individuals who have the same amount of strip searches, they would likely have different numbers of items found.

Table 5.1.4: ANCOVA analysis table between Items Found and (Perceived Race & Strip Searches)

Source	SS	DF	F	p-unc	np2
Perceived Race	876.620743	7	3.885961	7.586590e-04	0.186056
Strip Search	412826.839938	1	12810.103112	5.271731e-123	0.990796
Residual	3834.972562	119	N/A	N/A	N/A

To begin the discussion of the second ANCOVA analysis, we first note the null hypothesis: individuals from different perceived race groups have the same numbers of items found on average after accounting for their number of strip searches.

Statistically speaking, the p-unc stands for “uncorrected p-value”, which is less than 0.05. Thus, we can reject the null hypothesis that different perceived race groups result in the same average number of items found, even after controlling for the number of strip searches. Meaning that different perceived race groups have an impact on the average number of items found, after accounting for the number of strip searches.

Practical speaking, we hypothesized that the perceived race of individuals would be able to predict the average number of items found. From our results, we see that there is a statistically significant relationship between perceived race and the number of items found when controlling for the number of strip searches. This means that if we randomly select individuals from different race groups (e.g. “White” and “Latino) who have the same amount of strip searches, they would likely have different numbers of items found.

5.2 Logistic Regression Analysis and Findings

In this section of the report, we intend to present results and findings regarding our research questions 3. Specifically, we studied the binary outcome likelihood of being strip searched with respect to various explanatory variables sex, perceived race, age groups, and actions at arrest. We first established Chi-square Tests to evaluate whether each explanatory variable significantly impacts the outcome. Moreover, we also performed logistics regression analysis on the likelihood of being strip searched and interpreted our results using parameters such as odds ratios and confidence intervals. In addition, we evaluated the performance of our model using the accuracy score and confusion matrix.

5.2.1 Chi-square Tests

The following four tables serve the same purpose, so the results and findings will be given below all of them instead of each to avoid redundancy.

Table 5.2.1: Chi-square Test Results between Strip Search and Sex

test	lambda	chi2	dof	pval	cramer	power
pearson	1.000000	46.666778	1	8.414348e-12	0.026746	0.999999
cressie-read	0.666667	47.201657	1	6.404628e-12	0.026899	1.000000
log-likelihood	0.000000	48.318457	1	3.623258e-12	0.027215	1.000000
freeman-tukey	-0.500000	49.199202	1	2.312448e-12	0.027462	1.000000
mod-log-likelihood	-1.000000	50.118919	1	1.447051e-12	0.027718	1.000000
neyman	-2.000000	52.082980	1	5.320410e-13	0.028256	1.000000

Table 5.2.2: Chi-square Test Results between Strip Search and Perceived Race

test	lambda	chi2	dof	pval	cramer	power
pearson	1.000000	383.905230	7	6.731484e-79	0.076713	1.0
cressie-read	0.666667	393.557095	7	5.741896e-81	0.077672	1.0
log-likelihood	0.000000	415.961659	7	8.990462e-86	0.079852	1.0
freeman-tukey	-0.500000	435.866837	7	4.807727e-90	0.081740	1.0
mod-log-likelihood	-1.000000	458.869846	7	5.526993e-95	0.083870	1.0
neyman	-2.000000	516.162033	7	2.684502e-107	0.088951	1.0

Table 5.2.3: Chi-square Test Results between Strip Search and Age

test	lambda	chi2	dof	pval	cramer	power
pearson	1.000000	478.087546	8	3.526559e-98	0.085608	1.0
cressie-read	0.666667	519.658567	8	4.251214e-107	0.089252	1.0
log-likelihood	0.000000	655.155213	8	3.210908e-136	0.100215	1.0
freeman-tukey	-0.500000	850.391907	8	2.821159e-178	0.114175	1.0
mod-log-likelihood	-1.000000	1236.874393	8	1.032719e-261	0.137696	1.0
neyman	-2.000000	4172.591269	8	0.000000e+00	0.252908	1.0

Table 5.2.4: Chi-square Test Results between Strip Search and Actions at Arrest

test	lambda	chi2	dof	pval	cramer	power
pearson	1.000000	710.377999	4	1.972686e-152	0.104353	1.0
cressie-read	0.666667	633.468227	4	8.834586e-136	0.098542	1.0
log-likelihood	0.000000	519.253268	4	4.587799e-111	0.089217	1.0
freeman-tukey	-0.500000	458.622640	4	5.938467e-98	0.083847	1.0
mod-log-likelihood	-1.000000	413.091916	4	4.124246e-88	0.079576	1.0
neyman	-2.000000	353.012759	4	3.921637e-75	0.073562	1.0

The tables show the results of Chi-square goodness-of-fit tests for different test statistics, including lambda, Cressie-Read, log-likelihood, Freeman-Tukey, mod-log-likelihood, and Neyman. The degrees of freedom (DOF) are 1, 7, 8, and 4 for sex, perceived race, age, and actions at arrest respectively and the p-value for all the tests is less than a threshold $\alpha=0.05$, indicating strong evidence against the null hypothesis, which means explanatory variables sex, perceived race, age, and actions at arrest all have significant impacts on the outcome strip search.

Additionally, the tables provide measures of effect size such as Cramer's V and statistical power. These measures can help to interpret the practical significance of the test results. Overall, the results suggest that there is a significant difference between the observed data and the expected distribution.

5.2.2 Logistic Regerssion

Table 5.2.5: Logistics Regression Results

Dep. Variable	StripSearched	**No. Observations**	52188
Model	Logit	**Df Residuals**	52170
Method	MLE	**Df Model**	17
Dep. Variable	StripSearched	**No. Observations**	52188
Model	Logit	**Df Residuals**	52170
Method	MLE	**Df Model**	17
Date	Fri, 07 Apr 2023	**Pseudo R-squ.**	0.02892
Time	03:35:38	**Log-Likelihood**	-18507.
converged	True	**LL-Null**	-19058.
Covariance Type	nonrobust	**LLR p-value**	1.193e-223

	coef	**std err**	**z**	**P>|z|**	**[0.025**	**0.975]**
Intercept	-5.1861	0.357	-14.529	0.000	-5.886	-4.486
Sex[T.M]	0.2676	0.037	7.330	0.000	0.196	0.339
Perceived_Race[T.East/Southeast Asian]	-0.5921	0.067	-8.816	0.000	-0.724	-0.460
Perceived_Race[T.Indigenous]	0.1949	0.075	2.596	0.009	0.048	0.342
Perceived_Race[T.Latino]	-0.7779	0.107	-7.239	0.000	-0.988	-0.567
Perceived_Race[T.Middle-Eastern]	-0.7596	0.081	-9.375	0.000	-0.918	-0.601
Perceived_Race[T.South Asian]	-0.7871	0.078	-10.038	0.000	-0.941	-0.633

Perceived_Race[T.Unknown or Legacy]	-0.3010	0.058	-5.200	0.000	-0.414	-0.188
Perceived_Race[T.White]	-0.0121	0.033	-0.371	0.711	-0.076	0.052
Age[T.Aged 17 years and younger]	3.2972	0.363	9.088	0.000	2.586	4.008
Age[T.Aged 18 to 24 years]	3.1024	0.357	8.702	0.000	2.404	3.801
Age[T.Aged 25 to 34 years]	3.0378	0.356	8.539	0.000	2.340	3.735
Age[T.Aged 35 to 44 years]	2.9793	0.356	8.367	0.000	2.281	3.677
Age[T.Aged 45 to 54 years]	2.6907	0.357	7.530	0.000	1.990	3.391
Age[T.Aged 55 to 64 years]	2.4740	0.360	6.866	0.000	1.768	3.180
Age[T.Aged 65 and older]	2.0723	0.407	5.093	0.000	1.275	2.870
Age[T.Aged 65 years and older]	-0.2676	0.679	-0.394	0.694	-1.599	1.064
Actions_at_arrest	0.3172	0.023	13.651	0.000	0.272	0.363

Table 5.2.6: Odds Ratio Results

Intercept	0.005594
Sex[T.M]	1.306851
Perceived_Race[T.East/Southeast Asian]	0.553153
Perceived_Race[T.Indigenous]	1.215244
Perceived_Race[T.Latino]	0.459386
Perceived_Race[T.Middle-Eastern]	0.467845
Perceived_Race[T.South Asian]	0.455141

Perceived_Race[T.Unknown or Legacy]	0.740112
Perceived_Race[T.White]	0.987965
Age[T.Aged 17 years and younger]	27.036036
Age[T.Aged 18 to 24 years]	22.252105
Age[T.Aged 25 to 34 years]	20.858884
Age[T.Aged 35 to 44 years]	19.674831
Age[T.Aged 45 to 54 years]	14.741841
Age[T.Aged 55 to 64 years]	11.869735
Age[T.Aged 65 and older]	7.942914
Age[T.Aged 65 years and older]	0.765192
Actions_at_arrest	1.373303

We perform a logistic regression to examine the effects of sex, perceived race, age, and actions at arrest, on the likelihood that individuals are getting strip searched. None of the features are statistically significant. There are several significant results from Table 5.2.1 and Table 5.2.2: (1) The variable sex uses “Female” as the reference level, which is associated with the likelihood of an individual being arrested. The odds ratio is 1.306851 for male individuals, meaning that male is 1.31 times more likely to get strip searched, this result is statistically significant because of a p-value less than 0.05; (2) The variable perceived race uses “Black” as the reference level, it is associated with the likelihood of an individual being arrested. The odds ratio is statistically significant for all races except “White” which has a p-value of 0.711, meaning that compared to

Black individuals, White individuals do not have a significant difference in the likelihood of being strip searched. For the other races, “East/Southeast Asian” has the odds ratio of 0.553153, meaning that East/Southeast Asian individuals are 0.553153 times more likely to get strip searched than Black individuals (less likely compared to Black individuals), “Indigenous” has the odds ratio of 1.215244, meaning that Indigenous individuals are 1.215244 times more likely to get strip searched than Black individuals, “Latino” has the odds ratio of 0.459386, meaning that Latino individuals are 0.459386 times more likely to get strip searched than Black individuals (less likely compared to Black individuals), “Middle-Eastern” has the odds ratio of 0.467845, meaning that Middle-Eastern individuals are 0.467845 times more likely to get strip searched than Black individuals (less likely compared to Black individuals), “South Asian” has the odds ratio of 0.455141, meaning that South Asian individuals are 0.455141 times more likely to get strip searched than Black individuals (less likely compared to Black individuals), and “Unknown or Legacy” has the odds ratio of 0.740112, meaning that Unknown or Legacy individuals are 0.455141 times more likely to get strip searched than Black individuals (less likely compared to Black individuals); (3) The variable age uses “17 years and under” as reference level, it is associated with likelihood of an individual being arrested. The odds ratio is statistically significant for all age levels except “Aged 65 years and older” which has a p-value of 0.694, meaning that compared to individuals aged “17 years and under”, individuals “aged 65 and older” do not have a significant difference in the likelihood of being strip searched. For all the other age groups, the odds ratio for each age group decreases as the age in each group increases. Thus, even though this is a categorical variable, we can see this trend when an individual’s age increases, the likelihood of being strip searched decreases. (4) Increasing number of actions at arrest was associated with an increased likelihood of an individual getting strip searched. The odds ratio is 1.373303. For each additional action increase for the individual, the odds that the individual gets strip searched increases by 1.373303 times, this result is statistically significant.

Table 5.2.6: Odds Ratio Results

	Lower CI	Upper CI	OR
Intercept	0.002779	0.011260	0.005594
Sex[T.M]	1.216603	1.403795	1.306851
Perceived_Race[T.East/Southeast Asian]	0.484928	0.630976	0.553153
Perceived_Race[T.Indigenous]	1.048938	1.407916	1.215244
Perceived_Race[T.Latino]	0.372145	0.567077	0.459386
Perceived_Race[T.Middle-Eastern]	0.399147	0.548366	0.467845
Perceived_Race[T.South Asian]	0.390302	0.530751	0.455141
Perceived_Race[T.Unknown or Legacy]	0.660740	0.829018	0.740112
Perceived_Race[T.White]	0.926769	1.053201	0.987965
Age[T.Aged 17 years and younger]	13.277993	55.049526	27.036036
Age[T.Aged 18 to 24 years]	11.063606	44.755407	22.252105
Age[T.Aged 25 to 34 years]	10.386407	41.890624	20.858884
Age[T.Aged 35 to 44 years]	9.791090	39.535838	19.674831
Age[T.Aged 45 to 54 years]	7.318130	29.696367	14.741841

Age[T.Aged 55 to 64 years]	5.857956	24.051155	11.869735
Age[T.Aged 65 and older]	3.578156	17.631952	7.942914
Age[T.Aged 65 years and older]	0.202069	2.897618	0.765192
Actions_at_arrest	1.312157	1.437297	1.373303

Table 5.2.6 shows the 95% confidence intervals (CIs) and odds ratios (ORs) for different predictor variables in the logistic regression model.

For the intercept, the 95% CI is (0.002779, 0.011260), with an OR of 0.005594. This suggests that when all other predictor variables are equal to zero, the odds of the outcome variable (likelihood of being strip searched) are estimated to be between 0.28% and 1.13% with 95% confidence.

For the predictor variable "Sex[T.M]", the 95% CI is (1.216603, 1.403795), with an OR of 1.306851. This indicates that holding all other variables constant, the odds of the outcome variable for males are estimated to be between 1.22 and 1.40 times higher than for females with 95% confidence.

For the predictor variable "Perceived_Race[T.East/Southeast Asian]", the 95% CI is (0.484928, 0.630976), with an OR of 0.553153. This suggests that compared to the reference category “Black”, the odds of the outcome variable for individuals perceived as East/Southeast Asian are estimated to be between 44.7% and 51.5% lower with 95% confidence. The CIs for the other races have similar interpretations. However, notice that the race “White” has a 95% CI (0.926769, 1.053201), which contains 0 in the interval, meaning the result might not be statistically significant.

For the predictor variable "Age[T.Aged 17 years and younger]", the 95% CI is (13.277993, 55.049526), with an OR of 27.036036. This suggests that compared to the reference category “Aged 17 years and under”, the odds of the outcome variable for individuals aged 17 years and younger are estimated to be between 13.3 and 55.0 times higher with 95% confidence.

For the predictor variable "Actions_at_arrest", the 95% CI is (1.312157, 1.437297), with an OR of 1.373303. This suggests that, for each one-unit increase in this variable, the odds of the outcome variable increase by between 31.2% and 43.7% with 95% confidence.

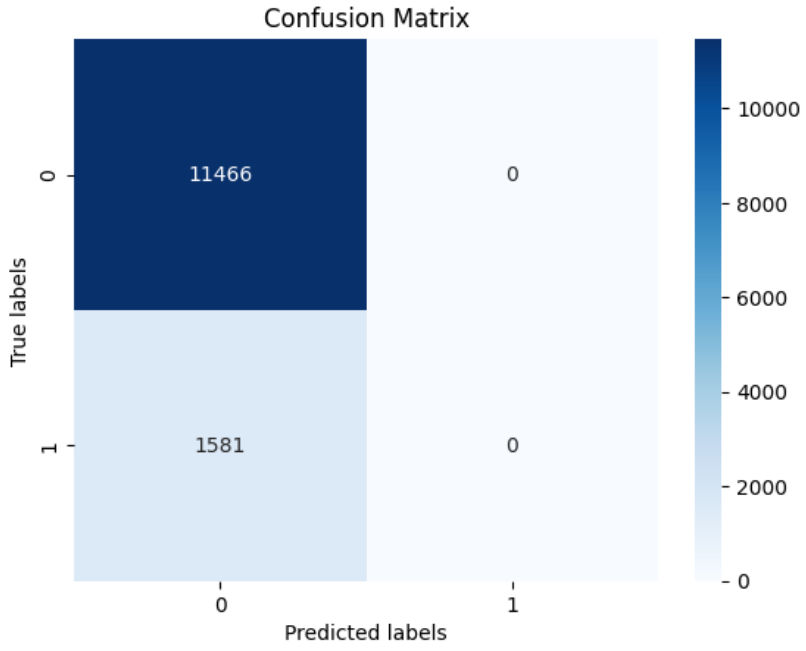
In conclusion, wider CIs indicate more uncertainty in the estimate of the true population parameter, while narrower CIs indicate more precision.

5.2.3 Test Accuracy and Confusion Matrix

After evaluating the test accuracy, we reached a score of 0.8788227178661762, meaning that the logistic regression model correctly predicted the class labels for approximately 87.88% of the test data. This suggests that the model is reasonably accurate in predicting the class labels for new data, which is encouraging.

Figure 5.2.1 visualizes the confusion matrix, it suggests that the logistic regression model predicted 11,466 instances as true positive (TP) and correctly identified them as negative, and predicted 1,581 instances as positive but incorrectly classified them as negative (false negatives, FN). However, the model did not predict any instances as positive (true negative, TN) or false positives (FP).

Figure 5.2.1: Confusion Matrix Result



Discussion and Conclusions

6.1 Limitations

Even if the test accuracy score is high, it's important to consider other performance metrics such as precision, recall, F1-score, and AUC-ROC score to get a more complete picture of the model's performance.

It's also important to keep in mind that the test accuracy score is just one measure of the model's performance and should be interpreted in conjunction with the context of the problem being solved. Depending on the problem and its consequences, a test accuracy score of 0.8788227178661762 may or may not be considered satisfactory. Therefore, it's important to define the performance threshold for the model beforehand and compare the test accuracy score against that threshold to determine whether the model is performing adequately.

6.2 Conclusion

In conclusion, this study provided an investigation of the arrests and strip searches dataset provided by the TPS Public Safety data portal. The team performed analysis and visualizations of the patterns and trends of the likelihood of being strip searched and the number of items found, following the research procedure of drawing research questions, data cleaning and preparation, descriptive analysis, and statistical analysis.

First of all, according to the power test, the significant result means that there is strong evidence to support the hypothesis being tested, suggesting a genuine relationship between the number of items found with sex and perceived race. After obtaining results from ANCOVA, it proves that different sex and different perceived race would likely have different numbers of items found at arrest. Furthermore, we performed a Chi-square test prior to fitting the Logistic regression model, and it reveals that sex, perceived race, age group, and the number of actions at arrest all have a significant impact on the likelihood of being strip searched. Lastly, the Logistic regression analysis shows a result that male individuals, Black individuals, individuals who are younger, and individuals who have more numbers of actions at arrest all have a higher chance of getting strip searched, Black individual has a higher likelihood of getting strip searched. The model accuracy reached approximately 87.88% with 0 false positives. Overall, this study successfully tested the hypothesis and answered the research questions. It is expected that it would provide insights on issues related to potential bias in the likelihood of being strip searched and items found at arrest in Toronto.

Reference

Arrests and strip searches (RBDC-arr-TBL-001). Toronto Police Service Public Safety Data Portal. (n.d.). Retrieved February 21, 2023, from https://data.torontopolice.on.ca/datasets/TorontoPS::arrests-and-strip-searches-rbdc-arr-tbl-001/about

Gills, W. (2019). *Toronto police conduct "far too many" strip searches, says damning report from Ontario's police watchdog*. Thestar.com. https://www.thestar.com/news/gta/2019/03/21/toronto-police-conduct-far-too-many-strip-searches-says-damning-report-from-ontarios-police-complaints-director.html

Gill, W. (2020, November 4). *"Clearly, we were doing it wrong": Toronto police are doing far fewer strip searches under strict new rules, Interim chief says*. Thestar.com. https://www.thestar.com/news/gta/2020/11/23/clearly-we-were-doing-it-wrong-toronto-police-doing-far-fewer-strip-searches-under-new-rules-interim-chief-says.html

Ontario Human Rights Commission. (2018). *Interim report on the inquiry into racial profiling and racial discrimination of Black persons by the Toronto Police Service.* https://www.ohrc.on.ca/sites/default/files/TPS%20Inquiry_Interim%20Report%20EN%20FINAL%20DESIGNED%20for%20remed_3_0.pdf#overlay-context=en/news_centre/ohrc-interim-report-toronto-police-service-inquiry-shows-disturbing-results

Toronto Police Service. (2021). *Chapter 1 -Arrest & Release*. https://www.tps.ca/media/filer_public/9b/d2/9bd29690-bfcf-42df-9480-7ba0d0ec6068/08da6409-baec-4b98-afe5-458c1e8d01b4.pdf

Toronto Police Service. (2020, June 25). *Race-Based Data Collection 2020 Findings - Strip Searches: Measurement & Outcomes (Video 5 of 6*). YouTube. https://www.thestar.com/news/gta/2020/11/23/clearly-we-were-doing-it-wrong-toronto-p https://www.youtube.com/watch?v=drSFau2QW7I