University of Toronto

INF2178 Final Project

Group Number: Group 50

By

Yijun Lin # 1003404620
Keying Zhu # 1003790953

INF 2178H S: Experimental Design for Data Science

Professor: Shion Guha

April 16th, 2023

Table of Content

Section 1: Introduction

1.1 Background Introduction

The Toronto Police Service (TPS) is the primary law enforcement agency responsible for policing the city of Toronto, Ontario, Canada. TPS has a clear mission to provide safety and security for Toronto residents and visitors while adhering to the principles of community policing, accountability, and professionalism. There is still room for improvement at TPS, and one of the most pressing issues facing TPS is racial profiling. Critics argue that even when there is no evidence of criminal activity, Toronto police disproportionately target BIPOC (black, Indigenous and people of colour) individuals, particularly Black and Indigenous individuals, for stops, searches, and arrests. This has resulted in a loss of trust between TPS and these communities, undermining the effectiveness of community policing and undermining public confidence in law enforcement. In recent years, TPS has come under increasing scrutiny and criticism over issues such as racial profiling, excessive use of force and lack of accountability.

This report seeks to conduct a comprehensive investigation into the association between the frequency of cooperative behaviour during arrests and the number of individuals subjected to strip searches by the Toronto Police Service (TPS). The study delves into the impact of demographic and socioeconomic variables, such as year, age, race, and sex, on the outcomes of these cooperative actions, while taking into account policy adjustments related to Covid-19. Furthermore, this report examines potential external factors that may affect an individual's propensity to cooperate during an arrest. Understanding the relationship between the prevalence of cooperative actions during arrests and the number of people subjected to strip searches, while adjusting for demographic and socioeconomic factors and acknowledging Covid-19 policy changes, is essential for devising recommendations to enhance the existing TPS system. By scrutinizing the available data, this report strives to deliver significant insights and suggestions that will contribute to the advancement of a more efficient and equitable TPS system. To offer such recommendations for the betterment of the TPS system, it is necessary to analyze the connection between law enforcement and the public. Consequently, this report will further investigate the relationship between TPS and the communities they serve, to provide additional recommendations for refining the current TPS system.

This report is structured into seven sections: Introduction, Tidy Data Procedure, Exploratory Data Analysis (EDA), Method, Results, Discussion, and Conclusion. Each section contributes to a comprehensive understanding of the dataset and its implications for the Toronto Police Service.

1.2 Literature Review

The principles of community policing, the impact of socioeconomic and demographic variables on interactions between the police and the public, and the implications of Covid-19 policy changes on law enforcement practices are the three main themes of the literature review for this report.

Community Policing and its Principles

Community policing has emerged as an essential strategy for law enforcement agencies to improve their relationship with the public and to enhance public safety (Weisburd & Eck, 2004). The core principles of community policing include problem-solving, partnerships, and organizational transformation (Oliver, 2016). Several studies have emphasized the significance of community policing in fostering trust between the police and the public, ultimately leading to increased public cooperation during arrests and other police-public interactions (Skogan & Frydl, 2004; Reisig et al., 2004).

Demographic and Socioeconomic Factors in Police-Public Interactions

The impact of demographic and socioeconomic factors on police-public interactions has been widely studied in the literature. Research has shown that factors such as age, race, and sex can influence an individual's likelihood of being subjected to a strip search, as well as their propensity to cooperate during an arrest (Gau & Brunson, 2010; Rios, 2011). Additionally, socioeconomic factors like income and education levels have been found to play a role in shaping police-public interactions (Papachristos et al., 2012). Understanding these factors is crucial for law enforcement agencies to develop equitable and effective policing strategies (Sampson & Lauritsen, 1997).

Covid-19 Policy Changes and Law Enforcement Practices

The Covid-19 pandemic has led to significant policy changes that have impacted law enforcement practices worldwide. These changes have resulted in law enforcement agencies adapting their procedures to protect public health and safety while maintaining effective policing (Lum et al., 2020). Studies have shown that the pandemic has influenced the frequency of certain police activities, such as arrests and searches, as well as shifting the focus toward enforcing public health-related regulations (Stickle & Felson, 2020). Understanding the implications of these policy changes is essential for evaluating the effectiveness of law enforcement agencies, like the Toronto Police Service, in the context of the pandemic.

1.3 Dataset Description

The data set "Arrests and Strip Searches (RBDC-ARR-TBL-001)" used in this report contains 65,276 data items collected and provided by the Toronto Police Service Public Safety Data Portal. The dataset contains information on 24 characteristics, including race, gender, age, and other demographic data, individuals arrested and strip-searched by TPS officers between January 1, 2020, and December 31, 2021(Toronto Police Service, 2022). The dataset includes information about each arrest and strip search, such as the reason for the arrest, where the individual was arrested, and the type of crime charged. The dataset also includes information on individuals' level of cooperation at the time of arrest, as well as the date and time of arrest and strip search. It is important to note that this dataset has scopes and limitations. Firstly, there are many missing values, meaning that some information may not be available for certain entries in the dataset. Secondly, there are too many categorical variables, which may make it more difficult to analyze the data. This dataset can be accessed via the Toronto Police Service Public Safety Data Portal: https://data.torontopolice.on.ca/datasets/TorontoPS::arrests-and-strip-searches-rbdc-arr-tbl-001/ and anyone have permission to view it (Toronto Police Service, 2022). The data is presented in various formats, including yes(1) or no(0) format, text-based format, and numerical integer format. The appendix section provides a full list of the attributes included in the dataset.

1.4 Research Questions/Objective

Based on the previous part that we discovered based on the current dataset and topic, our group picked the **main topic** as the **improvement of the current Toronto Police Service System**. At the same time, we analyzed one **overall research question** which is listed as followed:

1) How does the frequency of actions at arrest cooperative affect the number of people strip-searched, while controlling demographic and socioeconomic factors like the year(with the consideration of policies changed based on Covid-19), age, race and sex? If there exist any external factors that will influence the tendency of people for being cooperative actions at arrest?

Under this research approach, our group's **objective** is to **find out the improvement methods that can be taken by the Toronto Police Service System from the human interactive perspective** to see how the changing attitude/impression between policies and the public can improve the efficiency of the system operation with less probability of unequalized treatment.

Section 2: Tidy Data Procedure

2.1 Data Cleaning for Power Analysis & ANCOVA

Before the real application of the dataset, in order to better understand the related research topics, the data cleaning procedure is applied first. By reading the dataset into Python and checking the null value in each column, Table 1 shows the number of null values contained in each column which is listed below.

Column Name	**Null Value Counts**
Arrest_Year	0
Arrest_Month	0
Event ID	0

Arrest ID	469
Person ID	0
Perceived Race	4
Sex	0
Age_group_at_arrest	24
Younth_at_arrest_under_18_years	0
ArrestLocDiv	0
StripSearch	0
Booked	0
Occurrence Category	165
Actions_at_arrest_concealed_i	0
Actions_at_arrest_Combative_	0
Actions_at_arrest_Resisted_d	0
Actions_at_arrest_Mental_inst	0
Actions_at_arrest_Assaulted_o	0
Actions_at_arrest_Cooperative	0
SearchReason_CauseInjury	57475
SearchReason_AssistEscape	57475
SearchReason_PossessWeapons	57475
SearchReason_PossessEvidence	57475
ItemsFound	57475

Table 1 Counts of missing values for all columns

Based on Table 1 result, column Perceived_Race includes 4 null values which are relatively small values which might not highly affect the statistical result. The 4 missing values will be dropped from the original dataset. Second, column Age_group_at_arrest which is highly related which our research questions, this column contains 24 null values which is also a small amount compared to the overall number of rows which are also dropped from the original dataset. Also, columns relate to search reasons and ItemsFound since it contains too many null values, it is hard for us to make appropriate research questions and will make inaccurate statistical analysis results hence they will be removed from the dataset. Furthermore, since all columns related to ID as EventID, ArrestID, Personal ID and Object ID are used to count the number of rows as a way of dataset storage which is meaningless and hard to apply with the following procedures so those four columns are also dropped.

After dropping null values for selected columns, next the value counts for each column are applied to check whether the columns have the text format of null or unknown values and consider doing extra editions with that. By checking the value counted table based on Table 2, there are only two columns which need to take into consideration extra cleaning procedures. Column Perceived_Race contains the category of Unknown and Legacy with 5056 rows as the third largest categories in the current column of this dataset. Since it is reasonable to have situations when the polices cannot distinguish exact peoples' race and it has a relatively large share in race group, this category is kept. The Sex column consists of three categories, female, male and unisex, by comparing unisex with the other two categories, it only has 9 rows which makes the sample biased and might be hard to get statistical analysis results with other groups. Unisex is excluded from current research topics.

---- Perceived_Race ---	
Name	**Value counts**
White	27723
Black	17526
Unknown or Legacy	5056
East/Southeast Asian	4415
South Asian	3613
Middle-Eastern	3237
Indigenous	1934
Latino	1768
---- Sex ---	
Name	**Value counts**
M	52647

F	12626
U	9

Table 2 Patically value counted table for all columns

Mutation for categorical variables from string to numeric will help us consider the selection for variables in research questions based on correlation matrix results and form subset data frames. For some categorical variables that have similar concepts, such as “Aged 17 years and under” and “Aged 17 years and younger”, they will be mutated into the same number which is considered the same occurrence.

2.2 Data Cleaning for Logistic Regression

In order to capture better prediction and higher accuracy results, the data-cleaning procedure of logistic regression will primarily in consideration of converting null values into null first rather than dropping the null value in the first step. First, for columns related to search reasons, items found and all ID-related columns. They all have similar issues as in section 2.1, in considering the same reasons, our group dropped all those columns as they are uncorrelated with our research. Second, for column sex, we dropped the unisex rows with the consideration of too small samples. Column Age_group_at_arrest also has the merging steps to combine the same group with different word descriptions into one. Afterwards, by using the LabelEncoder and one Hot Encoder function, our group converts all categorial variables from the string format into the numerical format based on the types within each group which are mostly binary. At the same stage, the null value will be converted into 0 and for column ArrestLocDiv, in the data description from the website, “For some arrests, the location could not be geo-coded or the arrest took place outside of City of Toronto boundaries in other jurisdictions; these are indicated by XX”(Toronto Police Service, 2022) Referring to this, XX in this column can be considered as outlier which is replaced by a number of 99.

Section 3: Exploratory Data Analysis (EDA)

Since the application of the data cleaning procedure in section 2, the overview of data visualization for the current dataset will help us better understand which columns are suitable for the research questions and the approaches in creating subsets. A correlation matrix(Figure 1) shows the relationship between each variable as the majority of variables have relatively weak relationships with each other for all correlation coefficients are smaller than ±0.5. Reviewing back to the overall research question, StripSearch and Actions_at_arrest_Cooperative are two primary columns which will be considered as starting points for research. By checking their relationship with other variables in Figure 1, our group gets the following assumptions. First, factors Perceived_Race(-0.049), sex(-0.027) and Age_group_at_arrest_(-0.044) have relatively a weak negative relationship with binary variable strip-searching hence the arrest_year is added as an external factor for research question one as it has the highest negative correlation coefficient with a value of -0.31. For column Actions_at_arrest_Cooperative, all remaining variables are not highly correlated with it since they are all in correlation below ±0.1. Our ground finds that there does not exist a comparatively stronger relationship included for Actions_at_arrest_Cooperative and other variables, the selections for independent variables by considering Actions_at_arrest_Cooperative will rely on the general thinking with background introduction and literature review as supported materials.

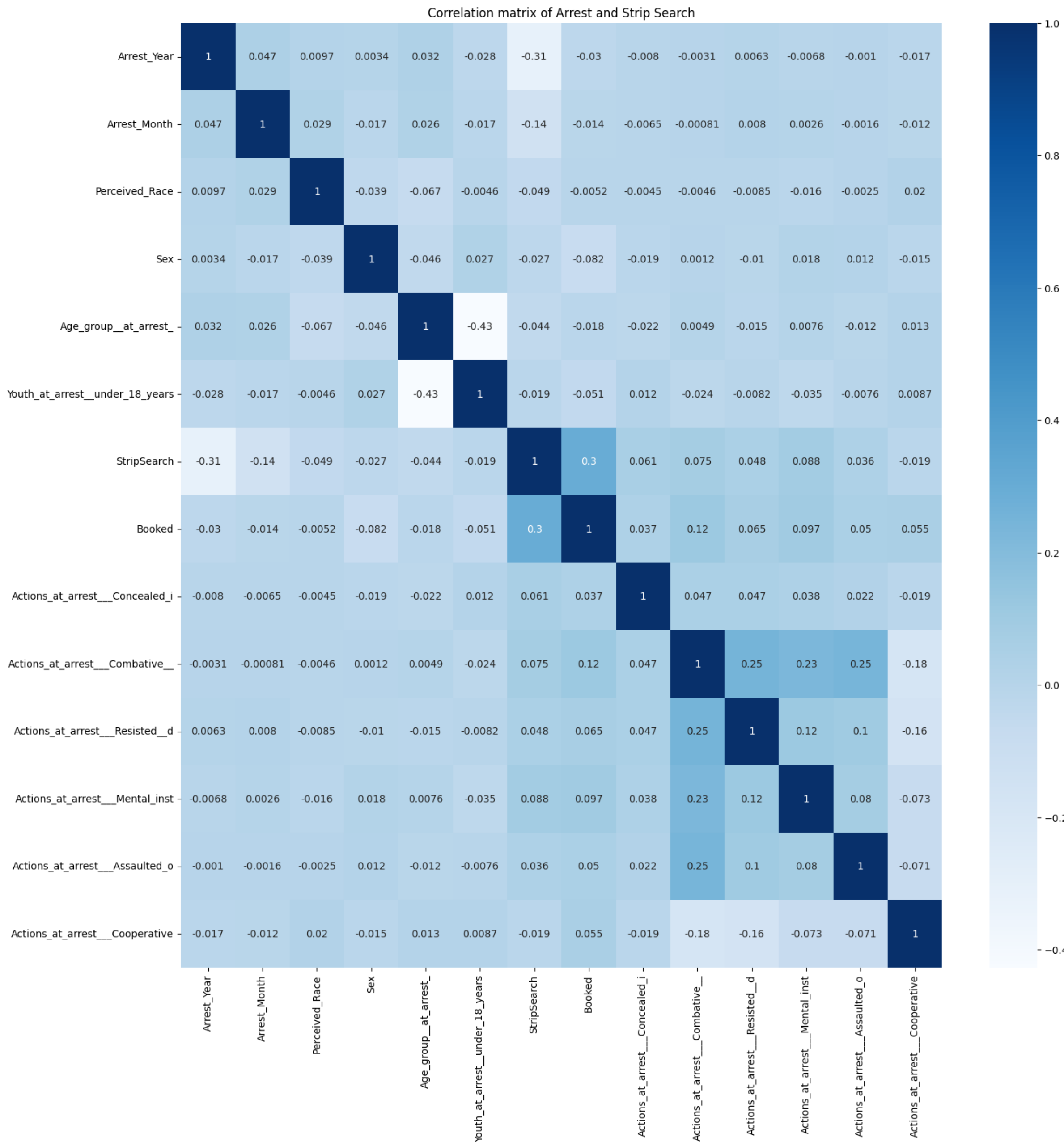

Figure 1 Correlation Matrix

Since the correlation matrix makes confirmation for data selections based on the overall research question, as the majority of columns from the dataset are either categorical or binary, this circumstance fails to satisfy the conditions to launch proper test statistics which require continuous variables for both dependent and independent. The data merging procedure in order to create continuous data columns is one of the main tasks for this section. The columns that are associated with the first part of the overall research question(How does the frequency of actions at arrest cooperative affect the number of people strip-searched while controlling demographic and socioeconomic factors like the year, age, race and sex?) Arrest_Year, Age_group_at_arrest, Perceived Race, Sex, Actions_at_arrest_Cooperative and StripSearch. For this question, our group aims to focus on the sub-dataset where people who are cooperative at arrest but were still being strip searched and those people are grouped based on different external factors such as year, age, sex and race. For Age_group_at_arrest, there exists some description for the age group which is logically considered as the same group which was already mentioned in section 2, those age groups age 17 years old and younger, age 65 and older will be replaced with the same string description to combine them as one group. Following this concept, firstly we need to filter out and count the frequency of people who are cooperative based on other independent factors then a similar procedure applies to people who are being strip searched. Finally, the two tables are merged together in order to fulfill both conditions. Hence Table 3 practically shows the finalized sub-dataset with 223 rows and 6 columns that will be used for the research question.

	Arrest_Year	**Perceived_ Race**	**Age_group_ at_arrest_**	**Sex**	**Cooperative_ Counts**	StripSearch_ Counts
0	2020	Black	Aged 17 years and younger	F	87	11
1	2020	Black	Aged 18 to 24 years	F	145	78
2	2020	Black	Aged 25 to 34 years	F	178	79
3	2020	Black	Aged 35 to 44 years	F	112	41
4	2020	Black	Aged 45 to 54 years	F	59	28
5	2020	Black	Aged 55 to 64 years	F	16	2
6	2020	Black	Aged 65 and older	F	5	0

7	2020	Black	Aged 17 years and younger	M	229	135
8	2020	Black	Aged 18 to 24 years	M	731	518
9	2020	Black	Aged 25 to 34 years	M	1119	728

Table 3 Finalized sub-dataset for the research question(Partially)

3.1 Descriptive Statistics

To be more familiar with the sub-dataset to help further research and understanding, we produce the following charts and tables. Figure 2 indicates the count of strip searches based on different 8 race types. It is unsurprising to see black and white races take the majority of strip search counts because compared to other races, white races took up the highest portion of Toronto's population and the black race was the group with racism issues/events in recent years which add the meaningfulness to this area of research. For other race groups, since they have relatively small counts to compare with, it is hard to see a clear trend or assumptions for this moment. Table 4 is the descriptive statistics for different race groups in consideration of counts by strip search and cooperative attitude which is another supportive evidence to indicate race differences. Under the equalized case group for all races, the black and white groups have the highest mean and standard deviation for both count types, representing the higher spread out.

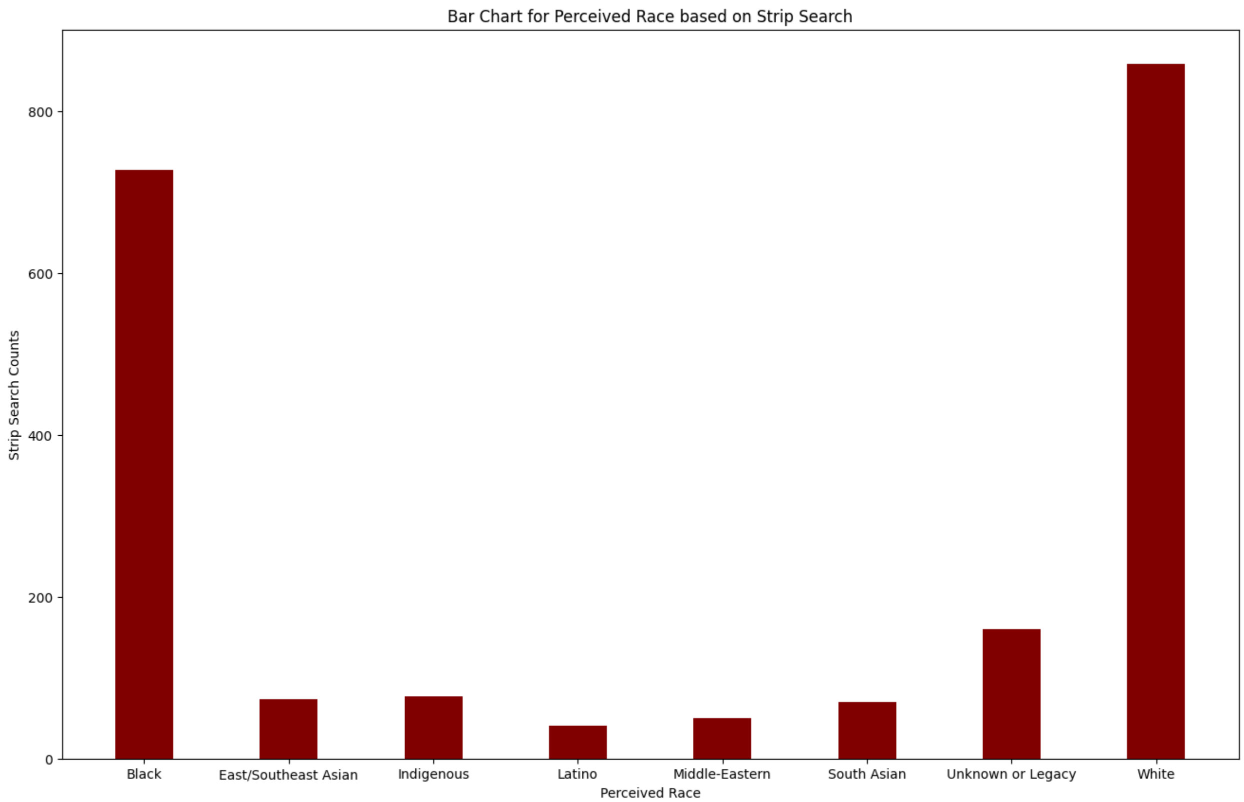

Figure 2 Bar Chart for Perceived Race based on Strip Search

	Strip Search Counts			**Cooperative at arrest Counts**		
	Count	**Mean**	**Standard Deviation**	**Count**	**Mean**	**Standard Deviation**
Perceived Race						
Black	28	86.93	172.63	28	266.57	316.61
East/SouthEast Asian	28	12.18	21.75	28	81.0	86.92
Indigenous	28	10.92	20.71	28	27.21	28.81
Latino	27	4.89	10.31	27	33.19	41.34
Middle-Eastern	28	8.14	15.39	28	54.07	65.57
South Asian	28	9.18	18.83	28	59.54	66.12
Unknown or	28	19.14	35.76	28	79.21	87.72

Legacy						
White	28	891.5	221.88	28	439.61	436.05

Table 4 Summary Statistics for Strip Search Counts and Cooperative Counts by race

By adding independent variable sex to form a boxplot in Figure 3, the majority race groups contain outliers. The white race group has the largest median for all genders. All female groups have lower strip search frequency than males which shows the differences between sex. For those 6 races except white and black, the female groups have less frequency of checking which shows as short boxplots. The scatter plot in Figure 4 examines the relationship between cooperative counts at arrest with counts of strip search based on year differences. For 2020 and 2021, cooperative counts and strip search counts have a positive relationship, as the number of cooperative counts increases, the number of counts for being strip searched increases. However, by comparing 2 years, the year 2020 has a stronger relationship for cooperative counts and strip search counts since it has a higher slope than the year 2021. At the same time, by using sex to find the relationship between cooperative counts at arrest and strip search counts for people, as the dots for both genders are majority centred at right corners with dispersed sharp. It is hard to analyze that there are no obvious gender differences for relationships in cooperative at arrest counts and strip search counts. Under the same gender, it is hard to see a clear relationship between cooperative counts and strip search counts.

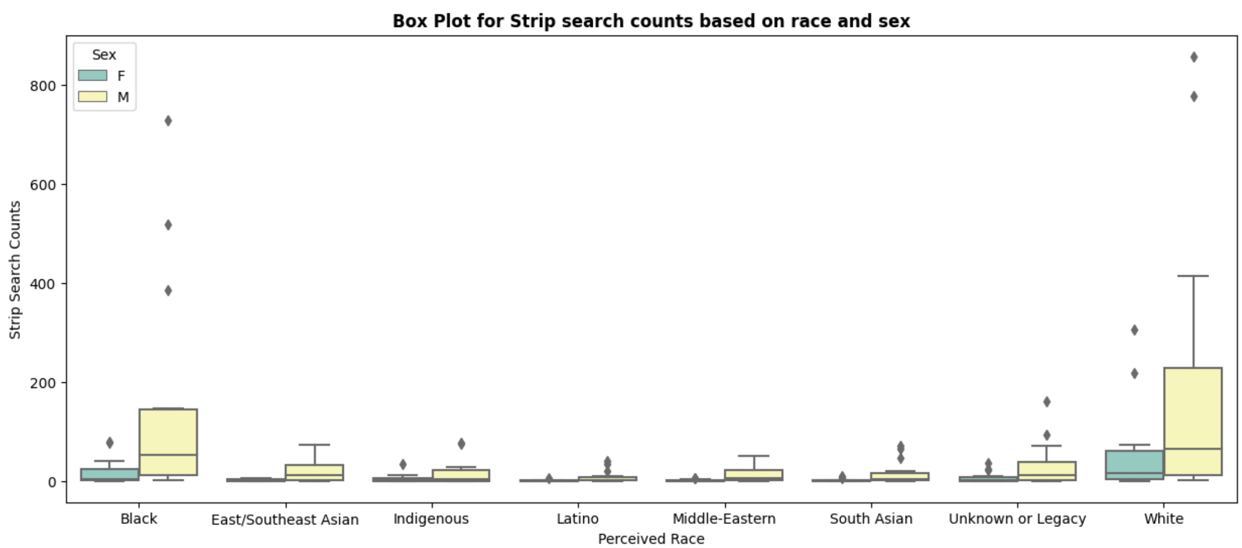

Figure 3 Boxplot for Strip Search count based on race and sex

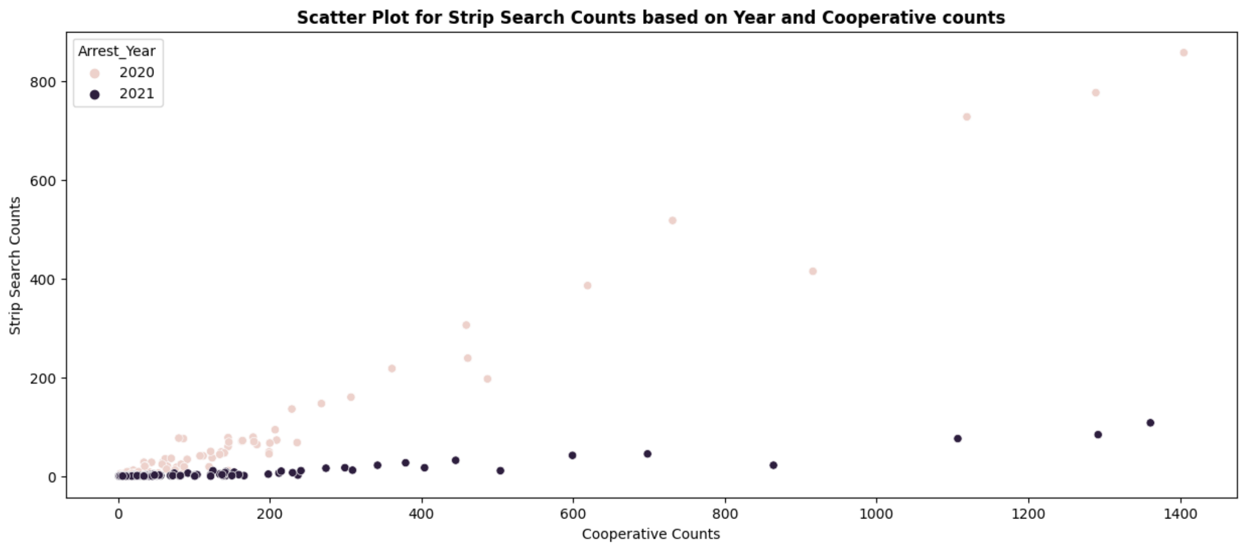

Figure 4 Scatter plot for Strip Search count based on years and Cooperative counts

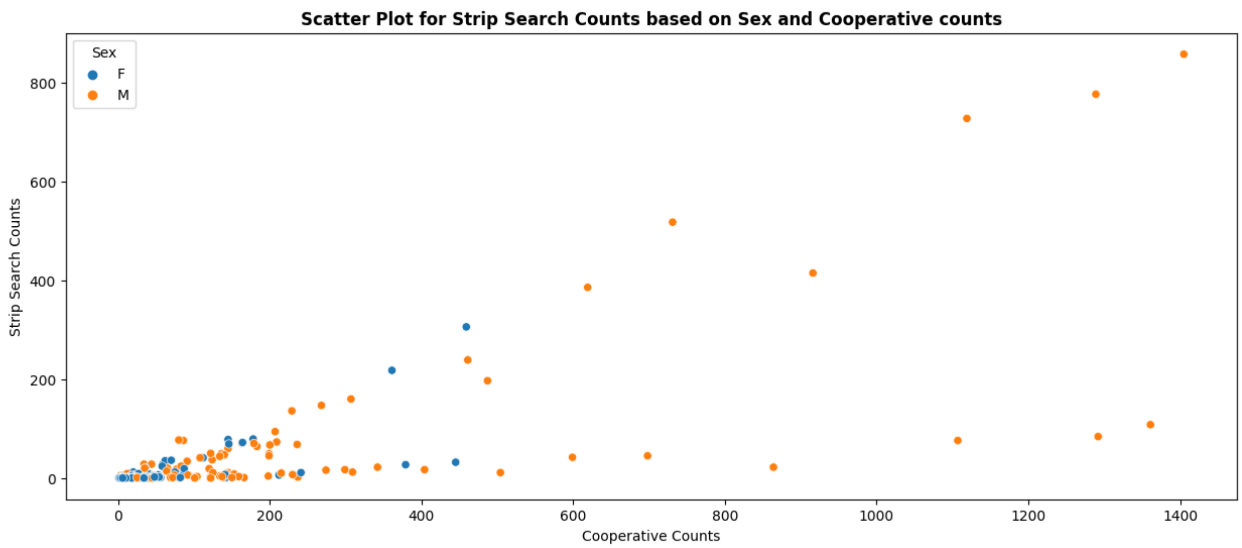

Figure 5 Scatter plot for Strip Search count based on Sex and Cooperative counts

As the independent variable Age_group_at_arrest contains 7 subgroups, the bar chart in Figure 6 indicates by controlling the person who has a cooperative attitude at arrest, how the count of being strip searched various under different age groups. Based on the bar chart, the young generation(Aged 25 to 34 years) and the middle-aged group(Aged 35 to 44 years) took the top 2 strip search counts. By checking the summary statistics for age groups based on two count types in Table 5, under majority equalized cases for each age group, the group Aged 25 to 34 years have the highest mean for both count types with the highest standard deviation. For cooperative at-arrest counts, the majority of age groups have a relatively high standard deviation to show the more spread out trending for those age groups. By adding sex as the previous procedure to the race group, the box plot based on age group and sex groups have outliers for the majority of groups except for the age group 65 and older. Similar to what bar chart and summary statistics table, aged 25 to 34 years and aged 35 to 44 years are the top two groups who have higher strip search counts by comparing with other groups. Also, excluding the age group 65 and older for visible consideration, the majority of male groups have higher strip search counts than females.

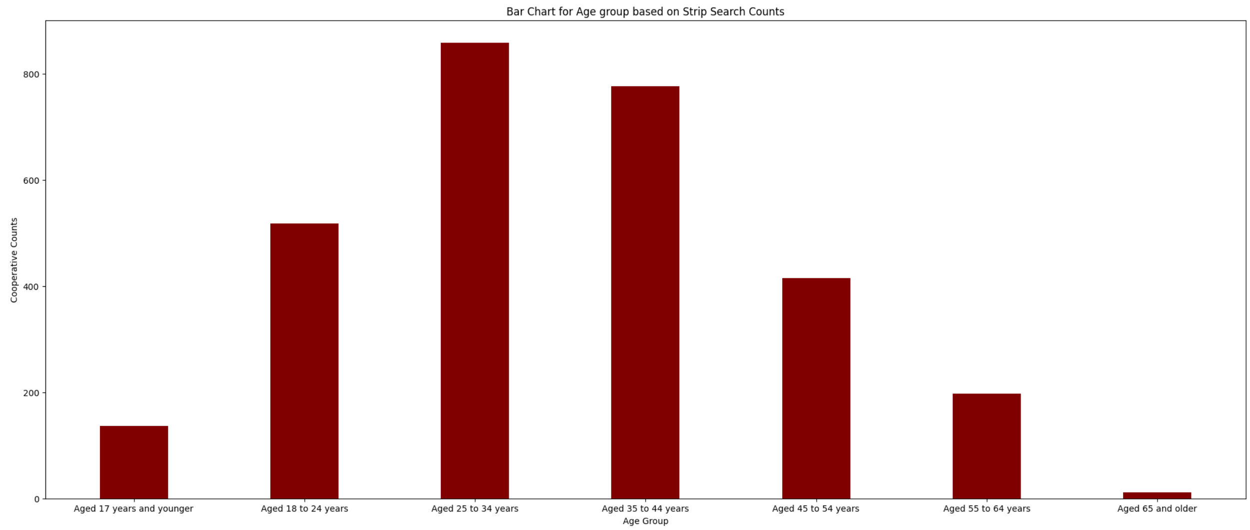

Figure 6 Bar Chart for Age group based on Strip Search

	Strip Search Counts			**Cooperative at arrest Counts**		
	Count	**Mean**	**Standard Deviation**	**Count**	**Mean**	**Standard Deviation**
Age Group						
Age 17 years and younger	32	8.75	24.87	32	44.25	62.16
Aged 18 to 24 years	32	42.16	98.16	32	138.06	173.72
Aged 25 to 34 years	32	85.59	195.71	32	285.34	384.90
Aged 35 to 44 years	32	65.53	76.31	32	225.47	331.88
Aged 45 to 54 years	32	28.28	36.22	32	130.78	215.42
Aged 55 to 64 years	31	11.68	2.49	31	67.35	120.66

Aged 65 and older	32	1.13	35.76	32	20.16	33.90

Table 5 Summary Statistics for Strip Search Counts and Cooperative Counts by age group

By adding sex as the previous procedure to the race group, the box plot based on age group and sex groups have outliers for the majority of groups except for the age group 65 and older. Similar to what bar chart and summary statistics table, aged 25 to 34 years and aged 35 to 44 years are the top two groups who have higher strip search counts by comparing with other groups. Also, excluding the age group 65 and older for visible consideration, the majority of male groups have higher strip search counts than females.

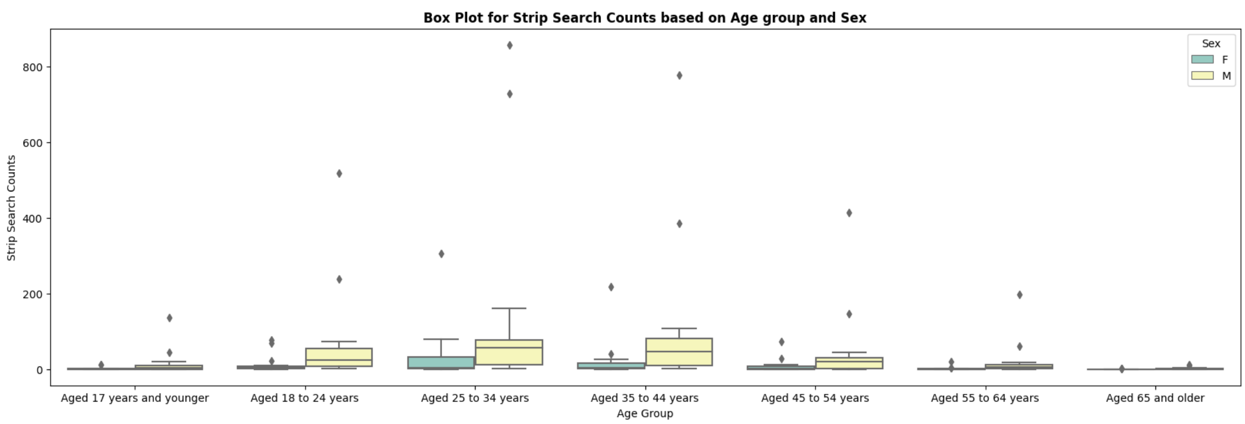

Figure 7 Boxplot for Strip Search count based on age group and sex

3.2 T-tests

Even though the procedure of mutation helps us understand the correlation between numerical variables with categorical variables, Welch's t-test is a more straightforward way to show relationships in different groups for categorical variables. We check the assumptions line by line for our sub-datasets and they are all satisfied: 1) nominal two-level explanatory variable; 2) quantitative variable as outcome variable; 3) normality and 4) independence of errors. For Welch's t-test we launch, there are no assumptions with equal variances among residuals. The paragraphs below demonstrate some useful statistical results extracted from the t-test.

Actual Size T-test

Sex and Strip Search Counts

For sex and strip search counts, we use Welch's T-test to analyze whether the population mean for strip search(outcome variable) actions will differ in sex groups(male and female as the explanatory variable). The hypothesis is stated as followed:

H_0(Null Hypothesis): For people who are cooperative at arrest, the population means for strip search counts for two independent groups: male and female are equal.

H_1(Alternative Hypothesis): For people who are cooperative at arrest, the population means for strip search counts for two independent groups: male and female are not equal.

By comparing with the significance level at 0.05, the t-test statistics result is -3.27 with a p-value of 0.012. As 0.012 < 0.05, there exist statistically significant differences. Therefore, we can reject the null hypothesis and conclude there are statistical differences between the male and female groups who are cooperative at arrest with the population mean for the strip search action.

Arrest Year and Strip Search Counts

For the year and strip search counts, we use Welch's T-test to analyze whether the population mean for strip search(outcome variable) actions will differ in different two years (2020 and 2021 as explanatory variables). The hypothesis is stated as followed:

H_0(Null Hypothesis): For people who are cooperative at arrest, the population means for strip search counts for two independent groups: 2020 and 2021 are equal.

H_1(Alternative Hypothesis): For people who are cooperative at arrest, the population means for strip search counts for two independent groups: 2020 and 2021 are not equal.

By comparing the significance level at 0.05, the t-test statistics result is 4.12 with a p-value of 5.37e-05. As 5.37e-05 < 0.05, there exist statistically significant differences. Therefore, we can reject the null hypothesis and conclude that for people who are cooperative at arrest, there are statistical differences between the year 2020 and year 2021 groups with the population mean for the strip search action.

Perceived Race and Strip Search Counts

For the year and strip search counts, we use Welch's T-test to analyze whether the population mean for strip search(outcome variable) actions will differ in two major race groups (white and non-white race groups as explanatory variables which is similar as). The hypothesis is stated as followed:

H_0(Null Hypothesis): For people who are cooperative at arrest, the population means for strip search counts for two independent groups: white and non-white race groups are equal.

H_1(Alternative Hypothesis): For people who are cooperative at arrest, the population means for strip search counts for two independent groups: white and non-white race groups are not equal.

By comparing the significance level at 0.05, the t-test statistics result is 5.07 with a p-value of 8.59e-07. As 8.59e-07 < 0.05, there exist statistically significant differences. Therefore, we can reject the null hypothesis and conclude that for people who are cooperative at arrest, there are statistical differences between the white race group and the non-white race group with the population mean for the strip search action.

Section 4: Methods

4.1 Power Analysis

Power is the true positive rate which can be defined as the probability of correctly rejecting a false null hypothesis from a statistical test. It also equals the 1 - Type 2 error where high type 2 error is generally worse than Type 1 error, the increase of power helps to reduce the probability of Type 2 error. Power with values of 0.8 is generally preferred and acceptable for a statistical test. Based on Seltman, there are several ways that can increase power such as increasing sample size, estimated effect size and significant level or reducing variability. To ensure adequate power for our t-tests, we applied sample size calculations based on effect size for each independent variable. Our group applies sample size and effect size for certain independent variables and here below are some statistical results and conclusions based on the selected independent variables. The sample size is calculated based on effect size as the sample size is needed for each sample group to achieve a given power for the t-tests. Adding a power plot helps visualize the relationship between effect size, sample size, and statistical power, making it easier to understand the study's statistical basis and enhancing the overall presentation of the research findings.

4.2 ANCOVA Tests

ANCOVA which is named analysis of covariance is “an analysis procedure for looking at group effects on a continuous outcome when some other continuous explanatory variable also has an effect on the outcome.”(Seltman, 2018) ANCOVA is used when blocking the continuous explanatory variable, to check how the simple mathematical relationship holds between the control variable and the outcome. Before the ANCOVA is applied to research questions, all assumptions that are listed are checked in detail: 1) Linearity; 2) Independence of observations; 3) Homoscedasticity; 4) Normality of residuals; 5) Additivity to make ANCOVA an appropriate method for our data analysis. By satisfying these assumptions, we can confidently apply ANCOVA to our dataset and obtain reliable and valid results, following paragraphs will list all the hypotheses that are made based on the selection of variables. Detailed interpretation, figures and results will be presented in Section 5.1.

Arrest Year, Cooperative at Arrest Counts and Strip Search Counts

For arrest year, cooperative at arrest counts and strip search counts, we used the ANCOVA test to examine the relationship between the arrest year differences and the number of people being strip-searched, while controlling for the cooperative actions at arrest as a covariate. The hypothesis is stated as followed:

H_0(Null Hypothesis):
There is no significant relationship between the arrest year and the number of people strip-searched when controlling the frequency of cooperative actions.

H_1(Alternative Hypothesis):
There is a significant relationship between arrest year and the number of people strip-searched while controlling the frequency of cooperative actions.

Sex, Cooperative at Arrest Counts and Strip Search Counts
For sex, cooperative at arrest counts, and strip search counts, we used the ANCOVA test to examine the relationship between the sex difference and the number of people being strip-searched, while controlling for the cooperative actions at arrest as a covariate. The hypotheses are stated as follows:

H_0(Null Hypothesis): There is no significant relationship between sex and the number of people strip-searched when controlling for the frequency of cooperative actions at arrest.

H_1(Alternative Hypothesis): There is a significant relationship between sex and the number of people strip-searched when controlling for the frequency of cooperative actions at arrest.

Perceived Race, Cooperative at Arrest Counts and Strip Search Counts
For perceived race, cooperative at arrest counts, and strip search counts, we used the ANCOVA test to examine the relationship between the different racial groups and the number of people being strip-searched, while controlling for the frequency of cooperative actions at arrest as a covariate. There are eight races in the dataset for comparison. The hypotheses are stated as follows:

H_0(Null Hypothesis): There is no significant relationship between perceived race and the number of people strip-searched when controlling for the frequency of cooperative actions at arrest.

H_1(Alternative Hypothesis): There is a significant relationship between perceived race and the number of people strip-searched when controlling for the frequency of cooperative actions at arrest.

Age group at arrest, Cooperative at Arrest Counts and Strip Search Counts

For age groups at arrest, cooperative at arrest counts, and strip search counts, we used the ANCOVA test to examine the relationship between the different age groups and the number of people being strip-searched, while controlling for the frequency of cooperative actions at arrest as a covariate. The hypotheses are stated as follows:

> H_0(Null Hypothesis): There is no significant relationship between the age group and the number of people strip-searched when controlling for the frequency of cooperative actions at arrest.
>
> H_1(Alternative Hypothesis): There is a significant relationship between the age group and the number of people strip-searched when controlling for the frequency of cooperative actions at arrest.

In conclusion, we employed the ANCOVA test to analyze the relationship between cooperative actions at arrest and the number of people strip-searched while controlling for various covariates, such as (a) Arrest year, to account for potential changes in policies and practices related to strip searches; (b) Sex, to explore potential differences in treatment based on gender; (c) Perceived race, to investigate potential disparities in treatment across different racial groups, with eight races included for comparison; (d) Age group at arrest, to examine the potential influence of age on the relationship between cooperative actions at arrest and the number of people strip-searched. By satisfying the assumptions of the ANCOVA test, we ensured that our analysis would yield reliable and valid results. The tables and results from these ANCOVA tests will be presented in Section 5.2 ANCOVA Test Results.

4.2 Logistic Regression

Logistic regression is a flexible method for modeling and testing the relationships between one or more quantitative and/or categorical explanatory variables and one binary (i.e., two-level) categorical outcome (Seltman, 2018). In the context of our research question, logistic regression will be used to which factor will influence the tendency of people to have a cooperative attitude at arrest. The detailed research question is in the format like this: Are there exist any personal factors such as race, age, year and arrest area with external factors such as year, booked or not and external actions at arrest which can influence the willingness for people to have a cooperative attitude at arrest. The reason why we keep other arrest actions into consideration is that there exist some cases in which people have multiple action records at arrest. Also, based on section 3.1, the correlation matrix indicates relatively weak correlation results between each action which shows there is no multicollinearity between them.

All assumptions that are listed are checked in detail: 1) The dependent variable is binary; 2) There is no multicollinearity among the independent variables; 3) The observations are independent of each other; 4) There is a linear relationship between the log odds of the dependent variable and the independent variables. By satisfying these assumptions, we can apply logistic regression to our dataset and obtain reliable and valid results.

The **purpose of logistic regression** provides a powerful tool for understanding how different factors contribute to the likelihood of an event or group membership. To validate our logistic regression model, we will employ a train-test split approach, which involves dividing the dataset into a training set and a testing set. The **training set** is used to build the model, while the **testing set** is used to evaluate its performance. This approach allows us to assess the model's ability to generalize to unseen data and provides an unbiased estimate of its predictive accuracy. **Confidence intervals** give us an estimate of the range within which the true population parameter (e.g., the regression coefficient) is likely to lie, while **prediction intervals** provide an estimate of the range within which the predicted probability of the binary outcome is likely to fall for a given set of predictor values. Both confidence and prediction intervals allow us to quantify the uncertainty associated with our estimates and make more informed decisions based on our analysis.

Section 5: Results/ Findings

5.1 Power Analysis Results

Based on the power analysis based on different categorical independent variables with controlling the continuous independent variable cooperative counts, here below stated the power analysis result for each variable:

Year on Strip Search Counts

Effect Size

The result of power analysis to calculate effect size is based on Cohen's D metric to see the effect size to analyze whether strip search counts(outcome variable) differed between year differences for people who are cooperative at arrest(two-level explanatory variable, the year 2020 and year 2021) which is 0.55. Based on the result, we can see effect size has a moderate impact on dependent variable strip search counts. The sample size that is required to estimate the mean number of people who are being booked for each year is calculated based on the assumption that the true population standard deviation equals the sample standard deviation.

Sample Size

After obtaining the effect size, the required sample size is computed using the obtained effect size and established based on 0.8 of statistical power. The results show that a sample size of 52 (rounded up by 52.32) was required for the year 2020 and a sample size of 53(rounded up by 52.79) was required for the year 2021. This is significant because the sample size provided in the dataset is 111 and 112 respectively, which impacts the reliability of the results. For both years, the calculated sample size results are smaller than the actual sample size results which will cause a less accurate estimation result for the means of strip search counts in the year 2020.

Sex on Strip Search Counts

Effect Size

The result of power analysis to calculate effect size is based on Cohen's D metric to see the effect size to analyze whether strip search counts(outcome variable) differed between different sex for people who are cooperative at arrest(two-level explanatory variable, male and female) which is 0.44. Based on the result, we can see effect size has a moderate impact on dependent variable strip search counts. The sample size that is required to estimate the mean number of people who are being booked for each year is calculated based on the assumptions as we made the independent variable year.

Sample Size

The results show that a sample size of 82(rounded up by 82.48) was required for the female group and a sample size of 83(rounded up by 83.20) was required for the male group. At the same time, the sample size provided in the dataset is 111 and 112 respectively which can impact the reliability of the results. For both genders, the calculated sample size results are smaller than the actual sample size results which will cause a less accurate estimation result for the means of strip search counts for males.

Perceived Race on Strip Search Counts

Review back to section 3.1, the white race group has the highest amount of strip search count with the highest mean and standard deviation. The race group will be divided into white and non-white groups(all of the remaining races) to analyze whether strip search counts(outcome variable) differed between two different race types for people who are cooperative at arrest (two-level explanatory variables, the white and non-white).

Effect Size

The result of power analysis to calculate effect size is based on Cohen's D metric to see the effect size to analyze whether strip search counts(outcome variable) differed between different sex for people who are cooperative at arrest(two-level explanatory variable, male and female) which is 1.02. Based on the result, we can see effect size has a high impact on dependent variable strip search counts. The sample size that is required to estimate the mean number of people who are being booked for each year is calculated based on the assumptions as we made the independent variable year.

Sample Size

The results show that a sample size of 9(rounded up by 8.81) was required for the white race group and a sample size of 61(rounded up by 61.38) was required for the non-white group. At the same time, the sample size provided in the dataset is 28 and 195 respectively which can impact the reliability of the results. For both race groups(white and non-white), the calculated sample size results are smaller than the actual sample size results which will cause a less accurate estimation result for the means of strip search counts for the non-white race group.

Power Plot Result

The power plot in Figure 8 shows the relationship between sample size and statistical power for three different effect sizes which are 0.2, 0.5 and 0.8. As we move along the sample size(x-axis), the power of the test (y-axis) increases with larger sample sizes. At the general significant level of 0.05, firstly this plot shows that at the effect size of 0.2, any sample size is hard to make the power of 80%. Since in our effect sizes results that were previously performed, there did not exist any effect size that is around 0.2 which is not our consideration or concern. If the effect size is 0.5, a sample of around 60 samples (over 50) is needed to achieve the power of 80%. Lastly, for the effect size at 0.8, a sample size of around 40 samples (below 50) is needed.

To sum up, the power plot indicates the sample size that is required to get the proper power result at 80% with the designated effect size. Based on the results we analyzed, for a smaller effect size to reach the power of 80%, a larger sample size is needed and vice versa for a larger effect size.

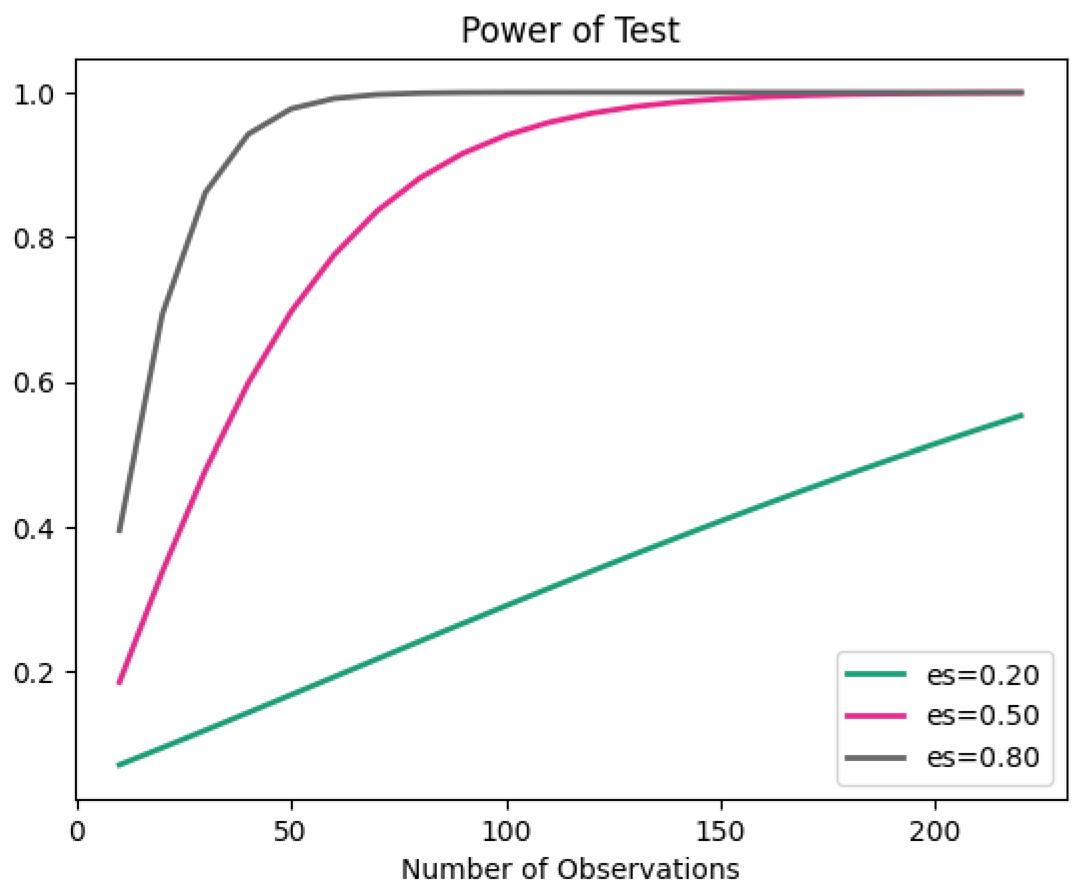

Figure 8 Power of Test for finalized sub-dataset

5.2 ANCOVA Test Results

Arrest Year, Cooperative at Arrest Counts and Strip Search Counts

The statistical results for ANCOVA for Arrest Year, Cooperative at Arrest Counts, and Strip Search Counts are shown in Table 6. By comparing the significance level at 0.05, our result indicates that the uncorrected p-value (0.00) is smaller than 0.05, revealing that there is a significant statistical relationship between arrest years and the number of people strip-searched while controlling for the frequency of cooperative actions at arrest. Thus, we reject the null hypothesis and conclude that each arrest year results in different strip search counts, after controlling for people's frequency of cooperative actions at arrest. The partial eta-squared for Arrest_Year (0.16) indicates that 16% of the total variability in the number of people strip-searched can be explained by the arrest year after controlling for the frequency of cooperative actions at arrest.

Source	SS	DF	F-values	Uncorrected p-values	Partial eta-squared
Arrest_Year	185282.91	1	40.93	0.00	0.16
Cooperative _counts	1443741.33	1	318.93	0.00	0.59
Residual	995876.01	220	NaN	NaN	NaN

Table 6 ANCOVA results for Arrest Year, Cooperative at Arrest Counts, and Strip Search Counts

Sex, Cooperative at Arrest Counts and Strip Search Counts

The statistical results for ANCOVA for Sex, Cooperative at Arrest Counts and Strip Search Counts are shown in Table 7. By comparing the significance level at 0.05, our result indicates that the uncorrected p-value(0.39) for Sex is greater than 0.05, revealing that there is no statistical relationship between sex and the number of people strip-searched when controlling for the frequency of cooperative actions at arrest. Thus, we fail to reject the null hypothesis that each gender results in the same strip search counts, after controlling for people's frequency of cooperative actions at arrest.

Source	SS	DF	F-values	Uncorrected p-values	Partial eta-squared
Sex	4039.68	1	0.76	0.39	0.00
Cooperative _counts	1328599.18	1	248.31	0.00	0.53
Residual	1177119.25	220	NaN	NaN	NaN

Table 7 ANCOVA results for Sex, Cooperative at Arrest Counts, and Strip Search Counts

Perceived Race, Cooperative at Arrest Counts and Strip Search Counts

The statistical results for ANCOVA for Perceived Race, Cooperative at Arrest Counts and Strip Search Counts are shown in Table 8. By comparing the significance level at 0.05, the uncorrected p-value(0.96) for Perceived Race is greater than the significance level, indicating that there is no significant relationship between perceived race and the number of people strip-searched when controlling for the frequency of cooperative actions at arrest. Thus, we fail to reject the null hypothesis that each race group results in the same strip search counts, after controlling for people's frequency of cooperative actions at arrest. This means

Source	**SS**	**DF**	**F-values**	**Uncorrected p-values**	**Partial eta-squared**
Perceived_Race	10672.58	7	0.28	0.96	0.01
Cooperative _counts	1040911.53	1	190.31	0.00	0.47
Residual	1177119.25	214	NaN	NaN	NaN

Table 8 ANCOVA results for Perceived Race, Cooperative at Arrest Counts, and Strip Search Counts

Age group at arrest, Cooperative at Arrest Counts and Strip Search Counts

The statistical results for ANCOVA for Age group at arrest, Cooperative at Arrest Counts, and Strip Search Counts are shown in Table 9. By comparing the significance level at 0.05, the uncorrected p-value(0.997) for the age group at arrest is greater than the significance level, indicating that after controlling for the age group at arrest, there is no significant relationship between the age group and the number of people strip-searched when controlling for the frequency of cooperative actions at arrest. Thus, we fail to reject the null hypothesis that each age group results in the same strip search counts, after controlling for people's frequency of cooperative actions at arrest.

Source	**SS**	**DF**	**F-values**	**Uncorrected p-values**	**Partial eta-squared**
Age_group_at _arrest	2988.87	6	0.091	0.997	0.00
Cooperative _counts	1255034.60	1	229.03	0.00	0.52
Residual	1178170.05	220	NaN	NaN	NaN

Table 9 ANCOVA results of ANCOVA for Age Group, Cooperative at Arrest Counts, and Strip Search Counts

5.3 Logistic Regression Results

The relationship between various predictor variables and the binary outcome variable "action at arrest___cooperation" is examined in the logistic regression results presented here. The dataset consists of 65267 observations and the model has 12 predictor variables which are arrest year, arrest month, perceived race, sex, age group at arrest, arrest location division, booked and 5 other actions/attitudes at arrest with the dependent binary variable cooperative actions at arrest. The model was estimated using Maximum Likelihood Estimation (MLE) and successfully converged. The coefficient (labelled coef in the following pages) of a predictor variable in a logistic regression model indicates the effect of that variable on the log odds of the outcome variable (in this case, cooperative behaviour during arrests). It provides an estimate of the direction and magnitude of the relationship between the predictor and the outcome. As shown in Table 10, our analysis revealed that several predictor variables had a statistically significant relationship with cooperative behaviour at arrest (p-value < 0.05).

	coef	**std err**	**z**	**P>|z|**	**[0.025**	**0.975]**
Intercept	111.8707	36.952	3.027	0.002	39.446	184.295
Arrest_Year	-0.0555	0.018	-3.035	0.002	-0.091	-0.020
ArrestLocDiv	-0.0002	0.000	-0.569	0.569	-0.001	0.000
Booked	0.3620	0.022	16.156	0.000	0.318	0.406
Actions_at_arrest___Concealed_i	-0.4104	0.170	-2.407	0.016	-0.745	-0.076
Actions_at_arrest___Combative__	-3.0100	0.121	-24.796	0.000	-3.248	-2.772
Actions_at_arrest___Resisted__d	-2.6864	0.112	-23.889	0.000	-2.907	-2.466
Actions_at_arrest___Mental_inst	-0.4553	0.062	-7.343	0.000	-0.577	-0.334
Actions_at_arrest___Assaulted_o	-4.4944	1.005	-4.471	0.000	-6.465	-2.524
Jan-Mar	0.0876	0.026	3.390	0.001	0.037	0.138
July-Sept	-0.00411	0.026	-0.158	0.875	-0.055	0.047
Oct-Dec	0.0206	0.026	0.777	0.437	-0.031	0.072
East/Southeast Asian	0.3346	0.039	8.540	0.000	0.258	0.411
Indigenous	-0.0261	0.058	-0.454	0.650	-0.139	0.087
Latino	0.3049	0.058	5.246	0.000	0.191	0.419

Middle Eastern	0.1844	0.044	4.170	0.000	0.098	0.271
South Asian	0.1543	0.042	3.647	0.000	0.071	0.237
Unkown Legacy	0.0279	0.037	0.755	0.450	-0.045	0.100
White	0.0740	0.023	3.202	0.001	0.029	0.119
M	0.0290	0.023	1.243	0.214	-0.017	0.075
Aged 18 to 24 years	-0.1273	0.048	-2.680	0.007	-0.220	-0.034
Aged 25 to 34 years	-0.1199	0.045	-2.692	0.007	-0.207	-0.033
Aged 35 to 45 years	-0.0988	0.046	-2.169	0.030	-0.188	-0.010
Aged 45 to 54 years	-0.0358	0.048	-0.742	0.458	-0.131	0.059
Aged 55 to 64 years	-0.0667	0.054	-1.235	0.217	-0.173	0.039
Aged 65 and older	0.0542	0.076	0.716	0.474	-0.094	0.203

Table 10 Summary Statistics results of Logistics Regression

Here below are the interpretation for all 25 independent variables with intercept terms with the dependent variable cooperative at arrest:

1) Intercept: If all the independent variables are with the value of 0, the log odds ratio for cooperative at arrest is 111.8707. By comparing with the significant level, it is statistically significant.
2) Arrest_Year: If one year/unit increases in arrest year, the log odds ratio for cooperative at arrest will decrease by 0.0555. By comparing with the significant level, it is statistically significant.
3) ArrestLocDiv: If one unit increases in arrest division(area changes), the log odds ratio for cooperative at arrest will decrease by 0.0002. By comparing with the significant level, it is not statistically significant.
4) Booked: If people are being booked, the log odds ratio for cooperative at arrest will increase by 0.3620. By comparing with the significant level, it is statistically significant.
5) Actions_at_arrest___Concealed_i: If people are having concealed item action at arrest, the log odds ratio for cooperative at arrest will decrease by 0.4104. By comparing with the significant level, it is statistically significant.
6) Actions_at_arrest___Combative__: If people are having combative action at arrest, the log odds ratio for cooperative at arrest will decrease by 3.01. By comparing with the significant level, it is statistically significant.
7) Actions_at_arrest___Resisted__d: If people are having resisted and defensive action at arrest, the log odds ratio for cooperative at arrest will decrease by 2.6864. By comparing

with the significant level, it is statistically significant.

8) Actions_at_arrest___Mental_inst: If people are having mental instability action at arrest, the log odds ratio for cooperative at arrest will decrease by 0.4553. By comparing with the significant level, it is statistically significant.
9) Actions_at_arrest___Assaulted_o: If people are having assaulted officer action at arrest, the log odds ratio for cooperative at arrest will decrease by 4.4944. By comparing with the significant level, it is statistically significant.
10) Jan-Mar: By comparing with arrest months Apr-June, if people are arrested in Jan-Mar, the log odds ratio for cooperative at arrest will increase by 0.0876. By comparing with the significant level, it is statistically significant.
11) July-Sept: By comparing with arrest months Apr-June, if people are arrested in July-Sept, the log odds ratio for cooperative at arrest will decrease by 0.00411. By comparing with the significant level, it is not statistically significant.
12) Oct-Dec: By comparing with arrest months Apr-June, if people are arrested in Oct-Dec, the log odds ratio for cooperative at arrest will increase by 0.0206. By comparing with the significant level, it is not statistically significant.
13) East/Southeast Asian: By comparing with the black race group, if people belong to East/Southeast Asian, the log odds ratio for cooperative at arrest will increase by 0.3346. By comparing with the significant level, it is statistically significant.
14) Indigenous: By comparing with the black race group, if people belong to Indigenous, the log odds ratio for cooperative at arrest will decrease by 0.0261. By comparing with the significant level, it is not statistically significant.
15) Latino: By comparing with the black race group, if people belong to Latino, the log odds ratio for cooperative at arrest will increase by 0.3049. By comparing with the significant level, it is statistically significant.
16) Middle Eastern: By comparing with the black race group, if people belong to Middle Eastern, the log odds ratio for cooperative at arrest will increase by 0.1844. By comparing with the significant level, it is statistically significant.
17) South Asian: By comparing with the black race group, if people belong to South Asian, the log odds ratio for cooperative at arrest will increase by 0.1543. By comparing with the significant level, it is statistically significant.
18) Unkown Legacy: By comparing with the black race group, if people belong to Unkown Legacy, the log odds ratio for cooperative at arrest will increase by 0.1543. By comparing with the significant level, it is not statistically significant.
19) White: By comparing with the black race group, if people belong to White, the log odds ratio for cooperative at arrest will increase by 0.0740. By comparing with the significant level, it is statistically significant.
20) M: By comparing with the female, if people are male, the log odds ratio for cooperative at arrest will increase by 0.0290. By comparing with the significant level, it is not statistically significant.

21) Aged 18 to 24 years: By comparing with the group Age 17 years and younger, if people are Aged 18 to 24 years, the log odds ratio for cooperative at arrest will decrease by 0.1273. By comparing with the significant level, it is statistically significant.
22) Aged 25 to 34 years: By comparing with the group Age 17 years and younger, if people are Aged 25 to 34 years, the log odds ratio for cooperative at arrest will decrease by 0.1199. By comparing with the significant level, it is statistically significant.
23) Aged 35 to 45 years: By comparing with the group Age 17 years and younger, if people are Aged 35 to 45 years, the log odds ratio for cooperative at arrest will decrease by 0.0988. By comparing with the significant level, it is statistically significant.
24) Aged 45 to 54 years: By comparing with the group Age 17 years and younger, if people are Aged 45 to 54 years, the log odds ratio for cooperative at arrest will decrease by 0.0358. By comparing with the significant level, it is not statistically significant.
25) Aged 55 to 64 years: By comparing with the group Age 17 years and younger, if people are Aged 55 to 64 years, the log odds ratio for cooperative at arrest will decrease by 0.0667. By comparing with the significant level, it is not statistically significant.
26) Aged 55 to 64 years: By comparing with the group Age 17 years and younger, if people are Aged 55 to 64 years, the log odds ratio for cooperative at arrest will increase by 0.0542. By comparing with the significant level, it is not statistically significant.

Table 11 presents the lower and upper limits of the 95% confidence intervals (CI) as well as the odds ratios (OR) for each predictor variable in the logistic regression model.

	Lower CI	**Upper CI**	**OR**
Intercept	1.352484e+17	1.092724e+80	3.844335e+48
ArrestYear	9.126931e-01	9.805265e-01	9.460020e-01
ArrestLocDiv	9.991678e-01	1.000458e+00	9.998128e-01
Booked	1.374512e+00	1.500703e+00	1.436222e+00
Actions_at_arrest___Concealed_i	4.749580e-01	9.266039e-01	6.633988e-01
Actions_at_arrest___Combative__	3.885439e-02	6.253118e-02	4.929108e-02
Actions_at_arrest___Resisted__d	5.464752e-02	8.492051e-02	6.812265e-02
Actions_at_arrest___Mental_inst	5.617062e-01	7.162426e-01	6.342853e-01
Actions_at_arrest___Assaulted_o	1.557763e-03	8.011955e-02	1.117172e-02
Jan-Mar	1.037648e+00	1.148275e+00	1.091561e+00

July-Sept	9.462074e-01	1.048192e+00	9.958949e-01
Oct-Dec	9.692005e-01	1.075076e+00	1.020766e+00
East/Southeast Asian	1.294111e+00	1.508955e+00	1.397410e+00
Indigenous	8.701687e-01	1.090671e+00	9.742011e-01
Latino	1.210442e+00	1.520128e+00	1.356476e+00
Middle Eastern	1.102685e+00	1.311437e+00	1.202539e+00
South Asian	1.073999e+00	1.267808e+00	1.166887e+00
Unknownor Legacy	9.563805e-01	1.105719e+00	1.028342e+00
White	1.029105e+00	1.126646e+00	1.076771e+00
M	9.834099e-01	1.077586e+00	1.029422e+00
Aged 18 to 24years	8.021811e-01	9.663875e-01	8.804645e-01
Aged 25 to 34years	8.129258e-01	9.679364e-01	8.870515e-01
Aged 35 to 44years	8.285733e-01	9.905153e-01	9.059330e-01
Aged 45 to 54years	8.776524e-01	1.060583e+00	9.647919e-01
Aged 55 to 64years	8.414479e-01	1.039936e+00	9.354422e-01
Aged 65 and older	9.100492e-01	1.224705e+00	1.055719e+00

Table 11 Summary Statistics Results of CI & Odds Ratio

An odds ratio of 1 indicates no effect, greater than 1 indicates increased odds of the outcome, and less than 1 indicates decreased odds of the outcome. Since the confidence intervals do not contain 1, we can conclude that the odds ratios are statistically significant at the 0.05 level. Here's an interpretation of the odds ratios for each predictor variable:

1) The odds ratio for the Intercept is 3.844335e+48, which suggests that the Intercept is not a meaningful predictor of the outcome variable.
2) The odds ratio for Arrest Year is 0.946002, which indicates that a one-unit increase in Arrest Year is associated with a 0.95 times decrease in the odds of the outcome variable.
3) The odds ratio for Booked is 1.436222, which indicates that individuals who were booked were 1.44 times more likely to have the outcome variable than those who were not booked, holding all other variables constant.

4) The odds ratio for Actions_at_arrest___Concealed_i is 0.663398, which indicates that individuals who concealed their actions during arrest were 0.66 times less likely to have the outcome variable than those who did not conceal their actions, holding all other variables constant.
5) The odds ratio for Actions_at_arrest___Combative__ is 0.049291, which indicates that individuals who were combative during arrest were 0.05 times less likely to have the outcome variable than those who were not combative, holding all other variables constant.
6) The odds ratio for Actions_at_arrest___Resisted__d is 0.068123, which indicates that individuals who resisted arrest were 0.07 times less likely to have the outcome variable than those who did not resist arrest, holding all other variables constant.
7) The odds ratio for Actions_at_arrest___Mental_inst is 0.634285, which indicates that individuals with mental instability during arrest were 0.63 times less likely to have the outcome variable than those without mental instability, holding all other variables constant.
8) The odds ratio for Actions_at_arrest___Assaulted_o is 0.011172, which indicates that individuals who assaulted officers during arrest were 0.01 times less likely to have the outcome variable than those who did not assault officers, holding all other variables constant.
9) The odds ratio for Jan-Mar is 1.091561, which indicates that events occurring between January and March are 1.09 times more likely to have the outcome variable compared to the Apr-June period, holding all other variables constant.
10) The odds ratio for July-Sept is 0.995895, which suggests that events occurring between July and September are not meaningfully different in terms of the outcome variable compared to the Apr-June period, holding all other variables constant.
11) The odds ratio for Oct-Dec is 1.020766, which indicates that events occurring between October and December are 1.02 times more likely to have the outcome variable compared to the Apr-June period, holding all other variables constant.
12) The odds ratio for East/Southeast Asian is 1.397410, which indicates that individuals of East/Southeast Asian ethnicity are 1.40 times more likely to have the outcome variable compared to the black race group, holding all other variables constant.
13) The odds ratio for Indigenous is 0.974201, which suggests that Indigenous individuals are not meaningfully different in terms of the outcome variable compared to the black race group, holding all other variables constant.
14) The odds ratio for Latino is 1.356476, which indicates that individuals of Latino ethnicity are 1.36 times more likely to have the outcome variable compared to the black race group, holding all other variables constant.
15) The odds ratio for Middle Eastern is 1.202539, which indicates that individuals of Middle Eastern ethnicity are 1.20 times more likely to have the outcome variable compared to the black race group, holding all other variables constant.

16) The odds ratio for South Asian is 1.166887, which indicates that individuals of South Asian ethnicity are 1.17 times more likely to have the outcome variable compared to the black race group, holding all other variables constant.
17) The odds ratio for Unknown or Legacy is 1.028342, which suggests that individuals of Unknown or Legacy ethnicity are not meaningfully different in terms of the outcome variable compared to the black race group, holding all other variables constant.
18) The odds ratio for White is 1.076771, which indicates that White individuals are 1.08 times more likely to have the outcome variable compared to the black race group, holding all other variables constant.
19) The odds ratio for Male is 1.029422, which suggests that males are not meaningfully different in terms of the outcome variable compared to the female group, holding all other variables constant.

20) The odds ratio for Aged 18 to 24 years is 0.880465, which indicates that individuals aged 18 to 24 years are 0.88 times less likely to have the outcome variable compared to the reference group, holding all other variables constant.
21) The odds ratio for Aged 25 to 34 years is 0.887051, which indicates that individuals aged 25 to 34 years are 0.89 times less likely to have the outcome variable compared to the Age 17 years and younger group, holding all other variables constant.
22) The odds ratio for Aged 35 to 44 years is 0.905933, which indicates that individuals aged 35 to 44 years are 0.91 times less likely to have the outcome variable compared to the Age 17 years and younger group, holding all other variables constant.
23) The odds ratio for Aged 45 to 54 years is 0.964792, which suggests that individuals aged 45 to 54 years are not meaningfully different in terms of the outcome variable compared to the Age 17 years and younger group, holding all other variables constant.
24) The odds ratio for Aged 55 to 64 years is 0.935442, which suggests that individuals aged 55 to 64 years are not meaningfully different in terms of the outcome variable compared to the Age 17 years and younger group, holding all other variables constant.
25) The odds ratio for Aged 65 and older is 1.055719, which indicates that individuals aged 65 and older are 1.06 times more likely to have the outcome variable compared to the Age 17 years and younger group, holding all other variables constant.

We could conclude that the following features are statistically significant: Booked, Actions at arrest (Concealed, Combative, Resisted, Mental inst, Assaulted officer), Ethnicity (East/Southeast Asian, Latino, Middle Eastern), and Age (18 to 24 years, 25 to 34 years, 35 to 44 years, 45 to 54 years, 55 to 64 years). The other features are not statistically significant.

The logistic regression model is split into trains and tests data with percentages of 80% and 20%. For each dataset, our group calculates the accuracy rate with the confusion matrix and makes comparisons between the two datasets. For the train dataset, the accuracy score is 0.5771, which means that the model correctly predicted cooperative behaviour during arrests in about 57.71% of the cases. The confusion matrix for the train dataset is as follows:

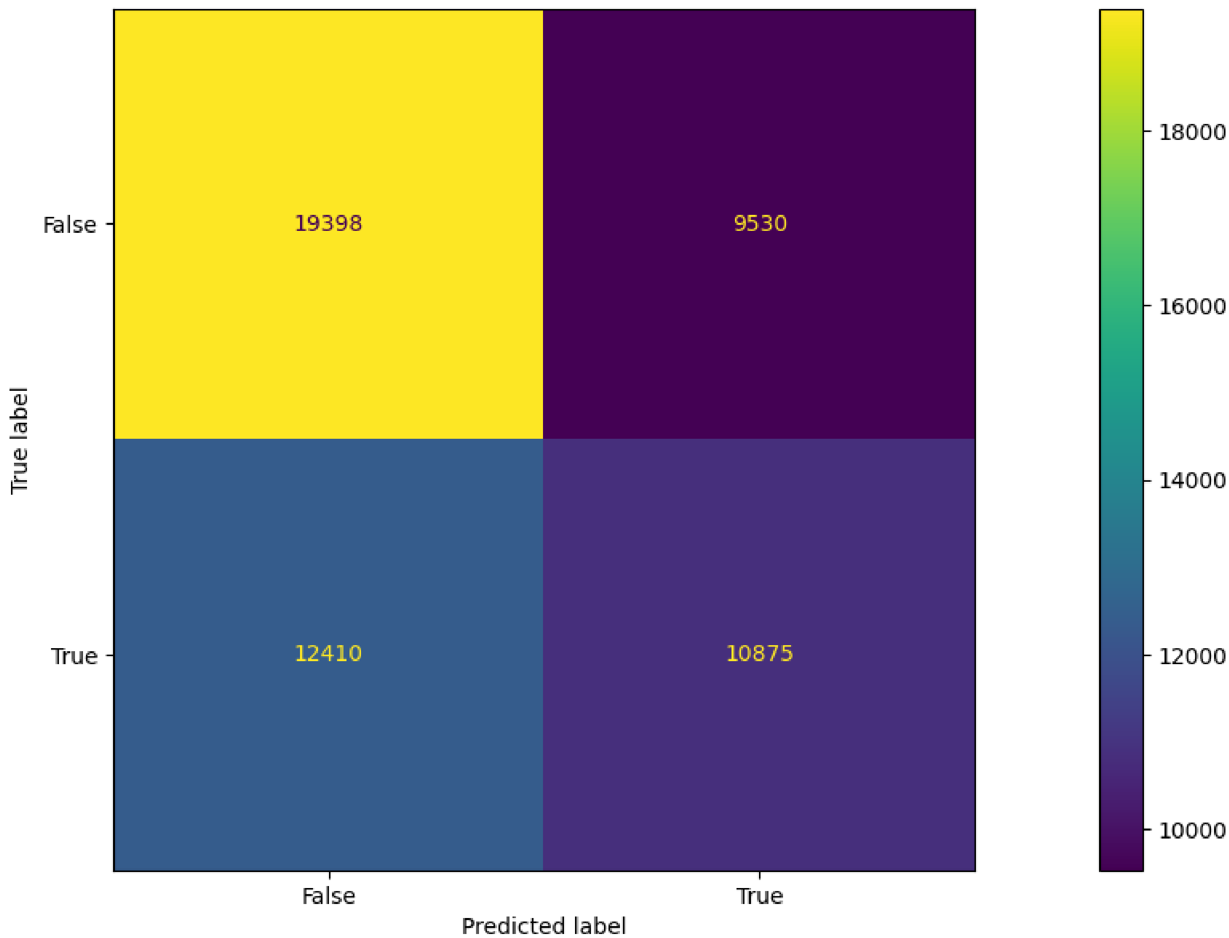

Figure 9 Confusion Matrix for Train dataset

In the train dataset, the model correctly predicted 19,398 true negatives (actual non-cooperative behavior predicted as non-cooperative) and 10,875 true positives (actual cooperative behavior predicted as cooperative). However, there were 9,530 false positives (actual non-cooperative behavior predicted as cooperative) and 12,410 false negatives (actual cooperative behavior predicted as non-cooperative).

For the test dataset, the accuracy score of 0.5810, which means that the model correctly predicted cooperative behavior during arrests in about 58.10% of the cases. The confusion matrix for the test dataset is as follows:

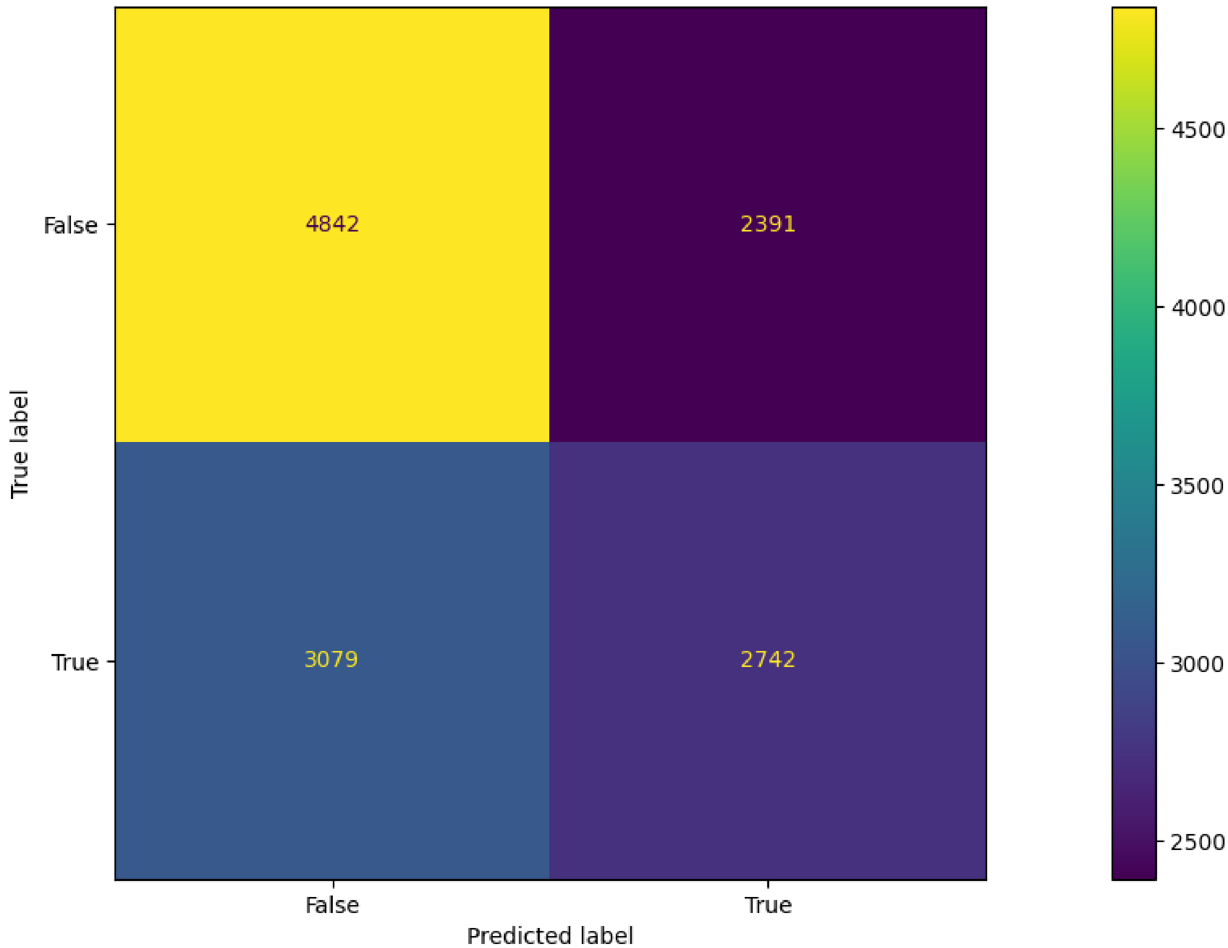

Figure 10 Confusion Matrix for Test dataset

In the test dataset, the model correctly predicted 4,842 true negatives (actual non-cooperative behavior predicted as non-cooperative) and 2,742 true positives (actual cooperative behavior predicted as cooperative). However, there were 2,391 false positives (actual non-cooperative behavior predicted as cooperative) and 3,079 false negatives (actual cooperative behavior predicted as non-cooperative).

For the train dataset, the accuracy score of 0.5798, which means that the model correctly predicted cooperative behavior during arrests in about 57.98% of the cases. The confusion matrix for the test dataset is as follows:

In general, it is not unusual for a model to have a slightly higher accuracy on the test dataset than the train dataset. Since in our case, the difference in accuracy between the training and test datasets is relatively small (0.5798 vs. 0.5810). This small difference may have no practical significance and may be due to random chance or for the reasons listed in Section 6 Discussion section.

The prediction interval in Table 12 is calculated based on the formula which shows the range of values that is likely to contain for each independent variable and intercept given the specific predictors. By comparing it with the confidence interval in Table 10, the prediction interval is slightly wider than the confidence intervals for the majority of independent variables which fit the concepts.

	Lower PI	**Upper PI**
Intercept	2.497539e-31	2.017859e+32
Arrest_Year	6.849112e+00	7.358153e+00
ArrestLocDiv	7.094492e+00	7.103654e+00
Booked	6.794046e+00	7.417791e+00
Actions_at_arrest___Concealed_i	5.082555e+00	9.915645e+00
Actions_at_arrest___Combative__	5.595943e+00	9.005956e+00
Actions_at_arrest___Resisted__d	5.694826e+00	8.849579e+00
Actions_at_arrest___Mental_inst	6.286748e+00	8.016357e+00
Actions_at_arrest___Assaulted_o	9.898810e-01	5.091199e+01
Jan-Mar	6.748444e+00	7.467916e+00
July-Sept	6.744882e+00	7.471860e+00
Oct-Dec	6.740449e+00	7.476774e+00
East/Southeast Asian	6.574294e+00	7.665738e+00
Indigenous	6.340980e+00	7.947796e+00
Latino	6.334808e+00	7.955540e+00
Middle Eastern	6.509588e+00	7.741936e+00
South Asian	6.533964e+00	7.713054e+00
Unkown Legacy	6.602289e+00	7.633234e+00
White	6.784809e+00	7.427890e+00

M	6.781767e+00	7.431222e+00
Aged 18 to 24 years	6.467882e+00	7.791857e+00
Aged 25 to 34 years	6.505843e+00	7.746393e+00
Aged 35 to 45 years	6.492866e+00	7.761875e+00
Aged 45 to 54 years	6.457887e+00	7.803917e+00
Aged 55 to 64 years	6.385748e+00	7.892077e+00
Aged 65 and older	6.119532e+00	8.235404e+00

Table 12 Statistical Results of Prediction Interval

Figure 11 and Figure 12 shows the prediction interval graph based on the selected independent variable perceived race and age group at arrest with dependent binary variable cooperative at arrest separately. Here below are the statistical results our group can extract from the prediction interval figures:

1) From the 95% of the predictive interval plot (as the selected significant level is at 0.05) in Figure 11, there exists an increasing trend with the positive relationship between different race groups and the probability of having cooperative actions at arrest, this model has a higher probability of making predictions with race groups white and unknown or legacy. However, for all probability for prediction interval, it is all below 0.5 which represents it is relatively low prediction accuracy and less reliable results.

2) From the 95% of the predictive interval plot in Figure 12, there exists an increasing trend with the positive relationship between different ages at arrest groups and the probability of having cooperative actions at arrest. As the people's age turns higher, the probability of prediction goes higher. However, with a similar issue that is faced for perceived race groups for all probability for prediction interval, it is all below 0.5 which represents it is relatively low prediction accuracy and less reliable results.

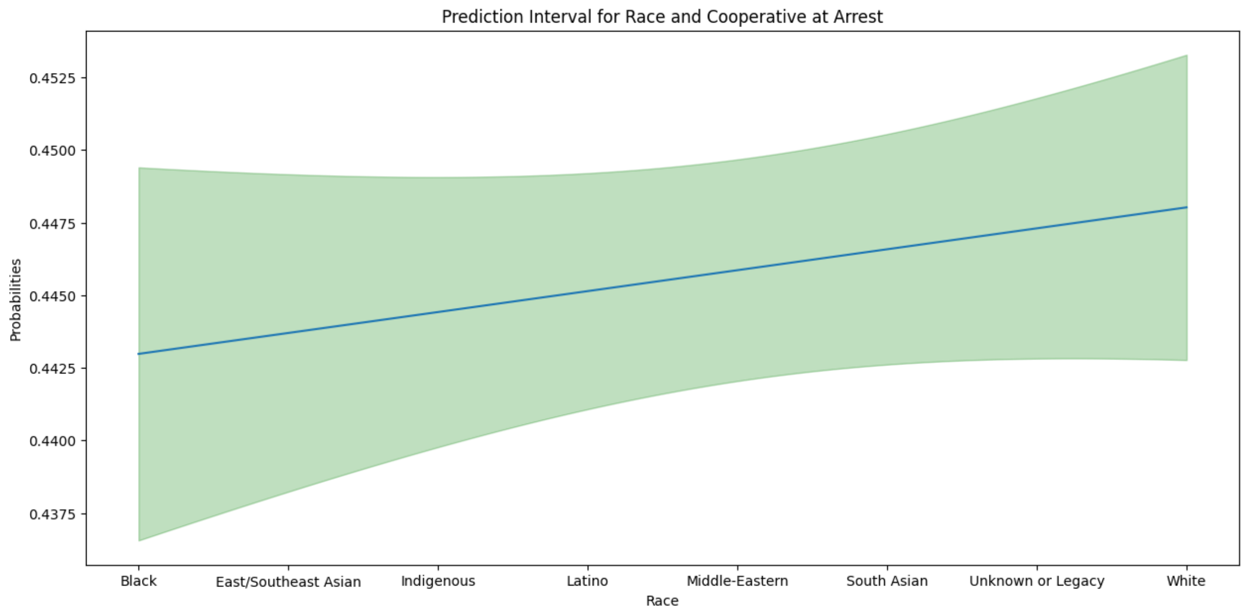

Figure 11 Prediction Interval for Race and Cooperative at Arrest

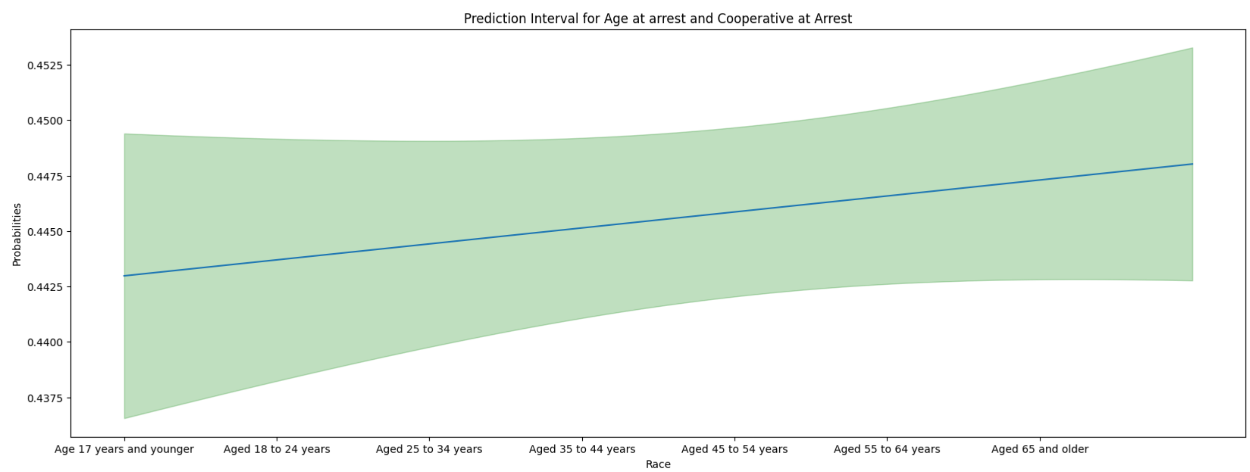

Figure 11 Prediction Interval for Age at Arrest and Cooperative at Arrest

Section 6: Discussion

6.1 Results Linkage

According to our results from previous sections within the literature review and overall research question, our group discover some points. Firstly, the power analysis (Section 5.1, page 22) indicated that the dataset used in this study has a sufficient sample size for the years 2020 and 2021, as well as for the male and female groups, thus providing reliable results for these

categories. The effect sizes obtained for these categories suggest that the independent variables have a moderate to large effect on strip search counts.
The ANCOVA test results (Section 5.2, pages 24) showed that the arrest year has a significant effect on the number of strip searches, after controlling for the frequency of cooperative actions at arrest. This suggests that there are differences in strip search counts between the years 2020 and 2021, which might be due to changes in law enforcement practices or policies during the COVID-19 pandemic.
However, no significant relationships were found between strip search counts and the variables of sex, perceived race, and age group at arrest(Section 5.2, pages 24-25), after controlling for cooperative actions. This indicates that within the context of this study, strip search practices do not disproportionately affect specific demographic groups.
According to our logistic regression analysis results (Section 5.3, pages 26-29), we observed that several predictor variables have a statistically significant relationship with cooperative behavior at arrest. These variables include booked, actions at arrest (concealed, combative, resisted, mental instability, assaulted officer), ethnicity (East/Southeast Asian, Latino, Middle Eastern), and age (18 to 24 years, 25 to 34 years, 35 to 44 years, 45 to 54 years, 55 to 64 years).

6.2 Limitations and Scopes

Firstly, the dataset contains numerous missing values, leading to potential gaps in the available information for some entries. This absence of data can introduce uncertainty and may affect the validity and generalizability of the findings derived from the analysis. Secondly, the dataset has an abundance of categorical variables, which can complicate the data analysis process. The handling and interpretation of categorical variables may require additional preprocessing steps and may present challenges in identifying meaningful relationships among the variables. While logistic regression can provide insights into the relationships between variables and the likelihood of a particular outcome, it cannot establish causality.

According to the statistical results for logistic regression in Table 10 on page 26, the moderate predictive accuracy of the model indicates potential avenues for improvement in aspects such as feature selection, engineering, or parameter tuning. To enhance the logistic regression's accuracy, variables with a p-value greater than 0.5, such as ArrestLocDiv, could be excluded from future analyses. Additionally, variables like ArrestLocDiv, Arrest_Month, and Year could be normalized to render the weights more comparable and facilitate a more accurate interpretation of the model's results.

According to the accuracy results and confusion matrix for the train and test dataset from page 33 to page 34, note that the model's accuracy is slightly higher on the test dataset than on the training dataset, a phenomenon that may seem counterintuitive but is not uncommon in machine learning. Several factors could contribute to this phenomenon. First, random chance could cause a difference in accuracy due to variations in the data, with the test dataset containing instances that the model can classify more easily even though it was not trained on them. Second, a relatively simple model that avoids overfitting may generalize better to unseen data, potentially improving performance on the test dataset. Third, differences in data distribution between the test and training sets can affect accuracy if the test set contains more instances from classes that the model can predict more accurately. Next, the smaller sample size of the test dataset can lead to a greater impact of a few correct predictions, resulting in higher test accuracy even if the model is not necessarily better at predicting test examples.

Finally, according to the logistic data selection, in our analysis, we encountered a large number of occurrence types and included many cases. Due to the wide range of these types, we had to impose certain restrictions and drop some of them to make the analysis more manageable and coherent. While this provides a comprehensive overview of the various cases, it also imposes some limitations due to the complexity of the dataset and a large number of categories. A large number of categories increases the potential for multicollinearity, making it difficult to determine the unique contribution of each occurrence type.

Section 7: Conclusion

Our study investigated the factors that influence the practice of strip searches in the Toronto Police Department and their impact on public safety and law enforcement-community relations. Consistent with previous research, our literature review highlights the importance of community policing principles in improving the relationship between law enforcement agencies and the public and in promoting cooperation during police-public interactions (Weisburd & Eck, 2004; Skogan & Frydl, 2004; Reisig et al.) Our findings also suggest that demographic and socioeconomic factors can play a role in influencing police interactions with the public and the likelihood of strip searches (Gau & Brunson, 2010; Rios, 2011; Papachristos et al., 2012). However, our study did not find a significant relationship between the number of strip searches and demographic factors (e.g., gender, perceived race, and age group at arrest) after controlling for cooperative actions.

The impact of COVID-19 policy changes on law enforcement practices is another important theme that emerged from our literature review (Lum et al., 2020; Stickle and Felson, 2020). Our ANCOVA test results suggest that differences in strip search counts between 2020 and 2021 may be due to changes in law enforcement practices or policies during the pandemic. The impact of these policy changes should be considered when assessing the effectiveness of law enforcement agencies (e.g., Toronto Police Service) with respect to strip-searching and policing.

Our research highlights the need to continue to investigate the factors that influence strip-search practices and their impact on public safety, law enforcement practices, and law enforcement community relations. We recommend that law enforcement agencies consider implementing community policing principles that take into account demographic and socioeconomic factors and adapt to policy changes resulting from the COVID-19 pandemic to improve their interactions with the public and promote fair and effective policing strategies.

The results of our analysis confirm previous research on cooperative behavior, and our results support the literature that emphasizes the importance of communication, empathy, and cultural competence in promoting cooperative behavior during arrests (Smith & Alpert, 2007; Engel et al., 2017) From a person-to-police perspective, these findings can guide law enforcement officers in developing targeted programs and training. By focusing on improving communication skills, increasing empathy, and developing cultural competence, law enforcement officers can better handle diverse situations and mitigate conflict. In addition, understanding how factors such as mental instability, gender, and behavior at the time of arrest affect cooperative behavior can help inform police tactics and protocols to ensure a more peaceful resolution. As suggested by Skogan (2006), building trust and cooperation between police and the community can lead to more effective and safer law enforcement practices. By addressing the factors identified in our study, policymakers can work to create a safer and more just society for all.

Google Colab link:
https://colab.research.google.com/drive/1upLqOsMS7zk_RUePP4fZBhGE0GW-qwEk?usp=sharing

Section 8: Reference

Engel, R. S., Sobol, J. J., & Worden, R. E. (2000). Further exploration of the demeanor hypothesis: The interaction effects of suspects' characteristics and demeanor on police behavior. *Justice Quarterly* https://doi.org/10.1080/07418820000096371

Gau, J. M., & Brunson, R. K. (2010). Procedural justice and order maintenance policing: A study of inner‐city young men's perceptions of police legitimacy. *Justice Quarterly*, 27(2), 255-279.

Lum, C., Maupin, C., & Stoltz, M. (2020). The impact of COVID-19 on law enforcement agencies (Wave 1). A joint report of the *International Association of Chiefs of Police and the Center for Evidence-Based Crime Policy*, George Mason University.

McCluskey, J. D., Terrill, W., & Paoline, E. A. (2005). Peer group aggressiveness and the use of coercion in police-suspect encounters. *Police Practice and Research*, 6(1), 19-37. https://doi.org/10.1080/1561426042000227647

Oliver, W. M. (2016). Community-Oriented Policing: A Systematic Approach to Policing. Pearson.

Papachristos, A. V., Meares, T. L., & Fagan, J. (2012). Why do criminals obey the law? The influence of legitimacy and social networks on active gun offenders. *Journal of Criminal Law and Criminology*, 102(2), 397-440.

Reisig, M. D., Bratton, J., & Gertz, M. G. (2007). The construct validity and refinement of process-based policing measures. *Criminal Justice and Behavior,* 34(8), 1005-1028.

Rios, V. M. (2011). Punished: Policing the Lives of Black and Latino Boys. *New York University Press.*

Sampson, R. J., & Lauritsen, J. L. (1997). Racial and ethnic disparities in crime and criminal justice in the United States. *Crime and justice*, 22, 311-374.

Seltman, H. J. (n.d.). *Carnegie Mellon Statistics & Data Science*. Retrieved April 8, 2023, from https://www.stat.cmu.edu/~hseltman/309/Book/Book.pdf.

Skogan, W. G., & Frydl, K. (2004). Fairness and Effectiveness in Policing: The Evidence. *National Academies Press*.

Stickle, B., & Felson, M. (2020). Crime rates in a pandemic: The largest criminological experiment in history. *American Journal of Criminal Justice*, 45(4), 525-536.

Toronto Police Service. (2022). *Arrests and strip searches (RBDC-arr-TBL-001)*. Toronto Police Service Public Safety Data Portal. Retrieved April 6, 2023, from https://data.torontopolice.on.ca/datasets/TorontoPS::arrests-and-strip-searches-rbdc-arr-tbl-001/About

Weisburd, D., & Eck, J. E. (2004). What can police do to reduce crime, disorder, and fear? *The Annals of the American Academy of Political and Social Science*, 593(1), 42-65.