INF2178 Final Submission
Comparative Analysis of Power Analysis, ANCOVA, and Logistic Regression on Criminal Stripe Search in Greater Toronto Area

Instructor: Shion Guha

Student:
Yajie Lu (1009236541)
Yang (Vera) Xuan (1003907427)

April 12, 2023

Table of Contents

Introduction

The report “Breaking The Golden Rule: A Review of the Police Strip Searches in Ontario" revealed that police officers in Ontario had performed around 22,000 strip searches annually, most of which were carried out by the Toronto Police Service (Write on Behalf of Barrison Law, 2021). Minority groups, particularly African Americans, were commonly overrepresented as offenders in the criminal justice system. As reported by the Uniform Crime Reports (UCR) in 2003, while only representing 12.7 percent of the total population, African Americans were arrested for 37 percent of violent crimes and 29 percent of property crimes. The situation meant that African Americans were more likely to be arrested for violent crimes, whereas whites were more likely to be arrested for burglaries and property crimes (ASA, 2007).

Despite most crimes being perpetrated by males, African American women are overrepresented within the criminal justice system. The rate of black women under the control of the criminal justice system is increasing faster than any other group, including black men and white men (ASA, 2007). The issue is whether the applicant's race or gender influenced the officer's failure to calm the situation. The power dynamic between a police officer and a member of the public inherently involves the exercise of power, given the authority granted to police officers by law. However, when this power dynamic is combined with racial and gender elements, it can be inappropriately amplified.

Discrimination based on race often stems from unconscious attitudes and beliefs, such as the belief that Black individuals should stay in their place and that those who resist should be punished severely. Based on these factors, the officer's actions appear consistent with racism, where a White person in authority expects obedience and punishes a person of color who does not comply (Tanovich, 2011). Intriguingly, recent arrest data from 1980 to 2010 shows a slight increase in the age of individuals arrested for crimes listed in the Uniform Crime Report (UCR) index in 2000 and 2010 compared to 1980. Additionally, there are differences in age patterns between specific offenses, and some offenses display more significant changes in age patterns over time than others. For instance, offenses such as robbery and aggravated assault exhibit little change in age patterns between 1980 and 2010, while offenses such as murder, rape, burglary, and auto theft show more noticeable variations (Ulmer & Steffensmeier, 2014).

Therefore, based on the research, we hypothesize that different age groups at arrest act the same, and no difference in the total number of strip searches conducted based on perceived race. Furthermore, we will conduct a power analysis to establish the necessary sample size for identifying the influence of various age groups on the probability of a strip search being conducted. Then we want to determine if the actions taken during the arrest significantly differ among the different age groups while accounting for the effects of occurrence category and perceived race. Finally, we will build a logistic regression model to predict the likelihood of a strip search across different perceived races. To facilitate our research, we will use the dataset from the Toronto Public Service, which includes information on age group at arrest, strip search, actions at arrest, occurrence category, and perceived race of the person for this project.

Literature Review

In the literature on social justice and crime, S*trip Searching* is defined as the "removal or rearrangement of some or all of someone's clothing" to allow police officers to inspect (McNeilly, 2019) visually. The righteousness of to which extent strip searching should be used in police inspection is subject to debate. In 2019, Toronto Police Service ("The Service") started to construct a "Race and Identity-based Data Collection Strategy" that is aimed at promoting the transparency of crime-related data and "repairing the trust of the community (Toronto Police Service, 2019)." Though the official source has reassured the public that the data collection is bias-free and inclusive, it is inevitable and hard to avoid inconsistencies associated with the data collection methods. The source of data is primarily reported and recorded by police officers. The report "Breaking the Golden Rule: A Review of the Police Strip Searches in Ontario" has revealed that race-related data are collected based on police officers' perceptions instead of self-reported by the individual associated with the strip search (McNeilly, 2019). Though the Service seeks improvement, it remains unclear whether the data collection would lead to bias or not – as there is yet no other systematic data collection solution.

Among existing studies, social scientists have argued that there is a disproportionate amount of police enforcement with armed or unarmed control being deployed to Black people (Toronto Police Service, 2022). Some scholars have questioned whether the official crime counting systems, commonly the data source for research, are intrusively biased (ASA, 2007). To counter systemic discrimination, the Service has rolled out a series of actions to address the potential of institutional discrimination (Global News 2023). The goals of the new approaches aimed to include the decision-making process of "using force or searching a person in situations that are unique, complex, and fluid (Toronto Police Service, 2022)." The Service adopted a cycle-based approach to replace the relatively linear process.

In this study, we aim to formulate our research inquiries through an extensive review of the relevant literature and subsequently employ statistical hypothesis testing to assess the validity and dependability of the Arrest and Strip Search dataset collected by the Toronto Police Services, with a particular emphasis on the fairness of the outcomes.

Research Objectives and Questions

Our study will examine age differences, racial groups, and the reason for being arrested interact with the chances of getting strip searched. Our analysis included an EDA analysis with motivation graphs to initiate aspects we want to identify the correlation between variables. We included a power analysis to identify the effective size of the age group being strip searched of the individual at the time of the arrest. We also conducted an ANCOVA test to examine how different age groups influence actions if one is arrested while controlling the occurrence reason as a control variable. Furthermore, with logistic regression, we predicted how perceived race affects the likelihood of being strip-searched during an arrest by using a machine learning model. Thus, we propose to research the following questions based on the preliminary research about the dataset (Arrest and Strip Searches (RBDC-AS-TBL-001) and the literature review on strip and arrest policy deployment disparity.

RQ1: What is the effect size of the individual's age group at the time of arrest on the likelihood of being strip-searched, and what sample size is needed to detect this effect with sufficient power?

RQ2: How does the age group of the individual at the time of arrest influence the actions taken during the arrest while controlling for the occurrence category of the arrest and the individual's perceived race?

RQ3: How does perceived race affect the likelihood of being strip-searched during an arrest, and what is the interpretation of the results, including the odds ratio?

We believe these questions will further facilitate our research on the occurrence and justice of the arrest and strip search cases in Greater Toronto.

Exploratory Data Analysis (EDA)

Dataset Description

In the proposed project, the dataset used in the analysis is named “Arrest and Strip Searches (RBDC-AS-TBL-001)”, which contains information related to all arrests and strip searches in the Great Toronto Area (Toronto Police Service, 2022). The data is gathered and published by the Toronto Police Service Public Safety Data Portal, an open-source website run by the Toronto Police Service platform, to improve public understanding of policing-related data and enhance information transparency (Toronto Police Service, 2022). The dataset collection is a vital part of the Toronto Police Service’s commitment to enhance the transparency of the “*Race and Identity-based Collection Strategy*,” highlighting some essential perspectives of each arrest case for researchers to navigate the dataset.

The table comprises 24 attributes that capture the information of arrested individuals and strip search-related information regarding each arrest case. In total, 65276 individual arrest records are stored in the dataset with a time range from 2020 to 2021. Notable demographic attributes include the perceived race, sex, and age group of each case. It is worth noticing that perceived race refers to the police officer’s perception of a person’s race instead of their self-identified race. The age groups of individuals being strip-searched or arrested are divided into nine groups to capture the character of combative actions after getting arrested or strip-searched.

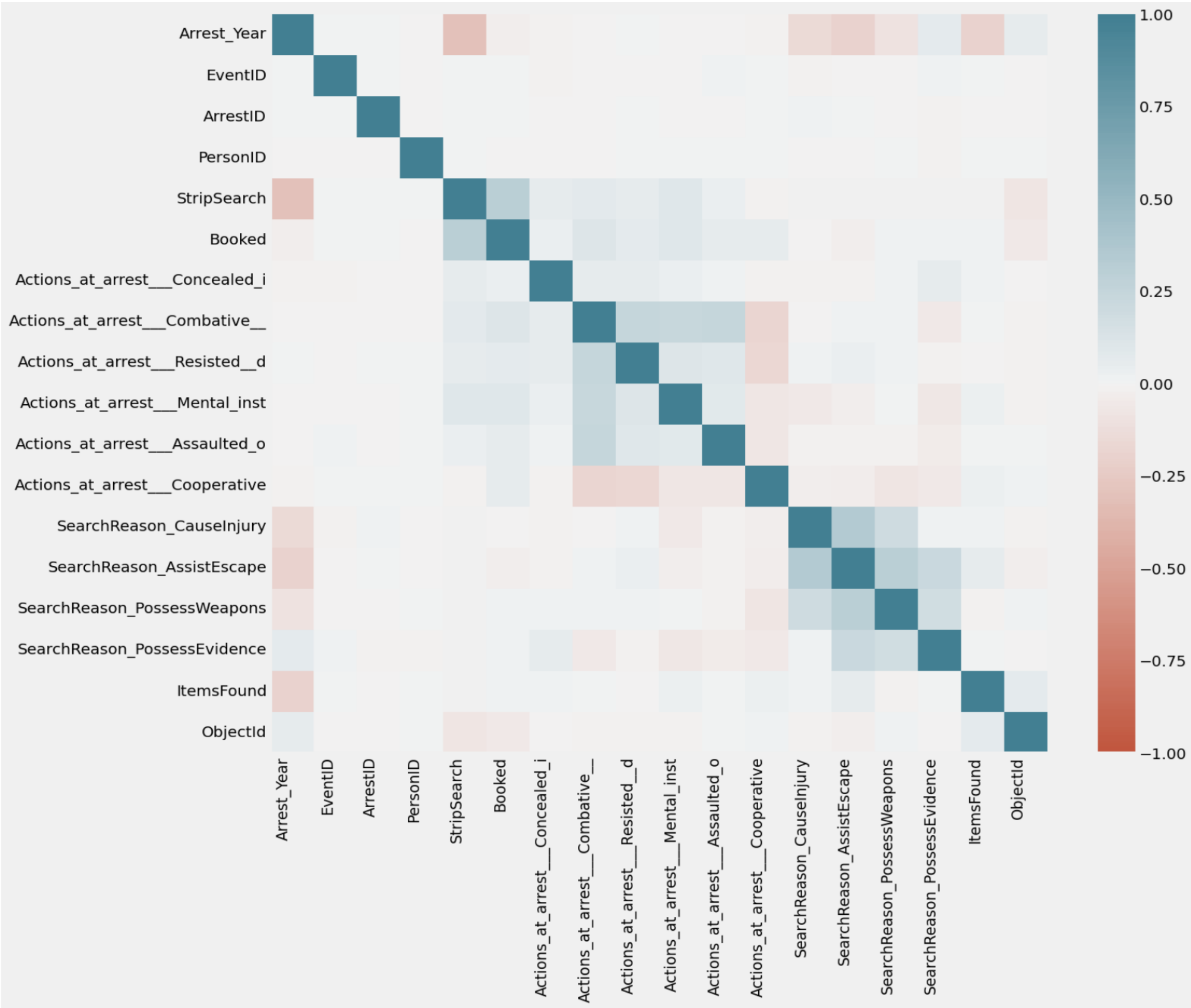
Specifically, in this research, we will mainly use sex, perceived_race, occurrence_category, Action_at_arrest, and Age_group to explore our research questions.

Figure 1 is a comprehensive interaction figure of all variables. After reading the .csv file into Python as a data frame, the shape of the dataset indicated that there are many missing values in different columns. The legend of the color bar demonstrated the correlation coefficient of each pair of variables. Using this graph, we can draw initial research motivations from the graph.

Descriptive Statistics & Motivation Graphs

Figure 1 is a comprehensive interaction figure of all variables. After reading the .csv file into Python as a data frame, the shape of the dataset indicated that there are many missing values in different columns. The legend of the color bar demonstrated the correlation coefficient of each pair of variables. Using this graph, we can draw initial research motivations from the graph.

Figure 1 Interaction Graph: All Variables



We started by locating missing values by incorporating a missing value matrix to explore the dataset further. Based on Figure 1, we can see that *ArrestID* and *Occurrence_Category* contain several missing cells, while the variables for the *SearchReason* and *ItemsFound* are missing majorities of data points. Thus, we used another verifying table (Figure 2) to create a more intuitive visualization of the bar chart to display the missing values. By observing the blue bar chart, we could confirm that *SearchReason* and *ItemsFound* contained the most missing values. To further explore, we created an additional graph that only contains the count of each search reason, as shown in Figure 3, to solidify the count for each category.

Figure 2 Visualization of Missing Values in the Dataset

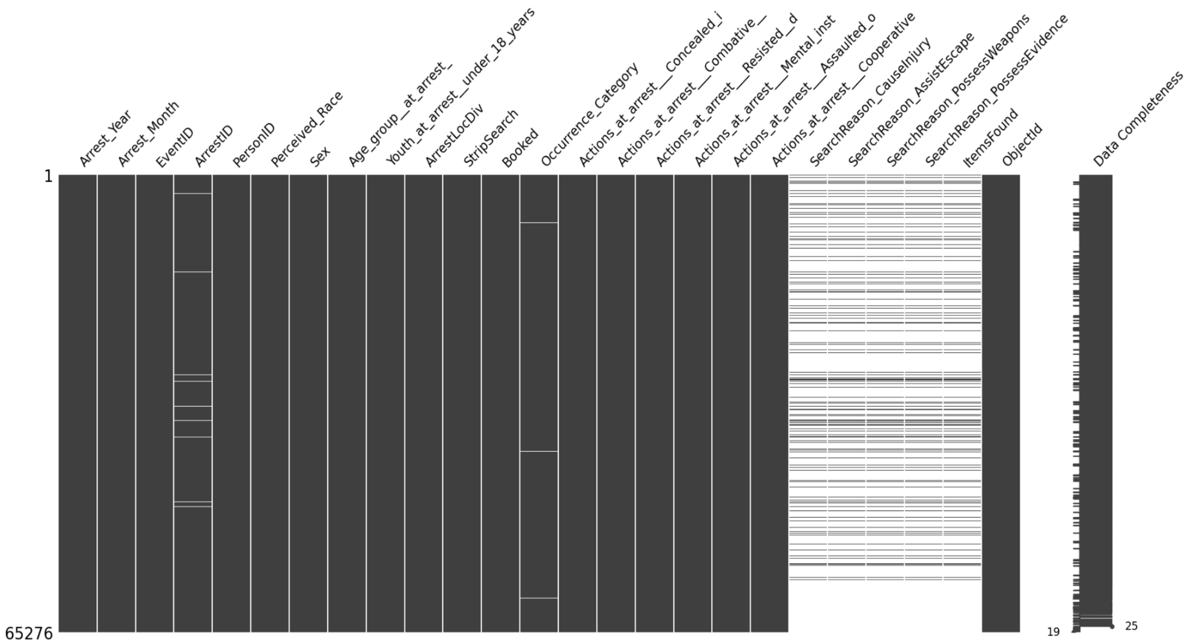
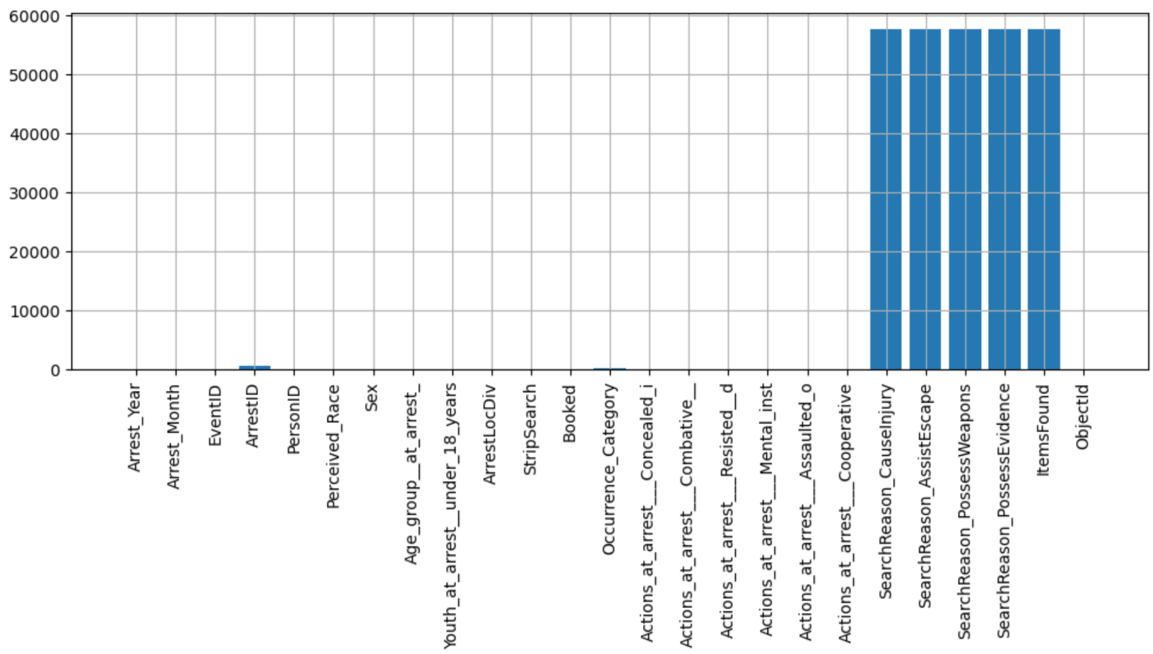


Figure 3 Visualization of Missing Values for ArrestID and SearchReasons



Since we intend to explore the impacts of a set of variables on whether a person is strip-searched, Figure 4 presents the distribution of stripe search counts where 0 represents those who were not strip-searched and 1 represents those who were strip-searched. Next, we explored the count of the Occurrence Category with several visualizations to use as a motivation graph for our research questions. Figure 5 are bar charts that we created for the variable of *Occurrence_Category*, capturing the counts of each criminal incurrence. From Figure 5, we can see which categories have the highest to the lowest counts.

Figure 4 Visualization of Strip Search Counts

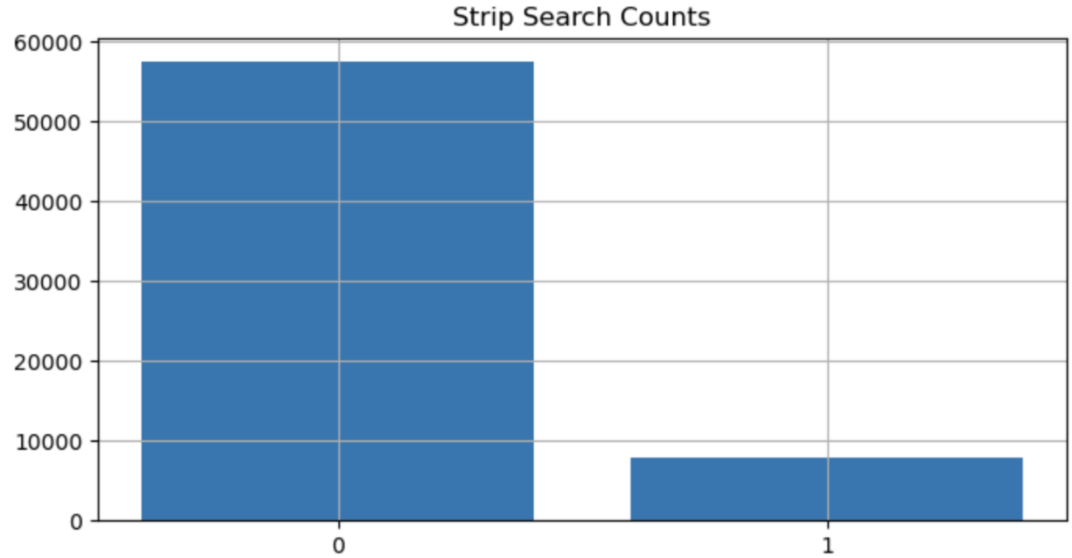
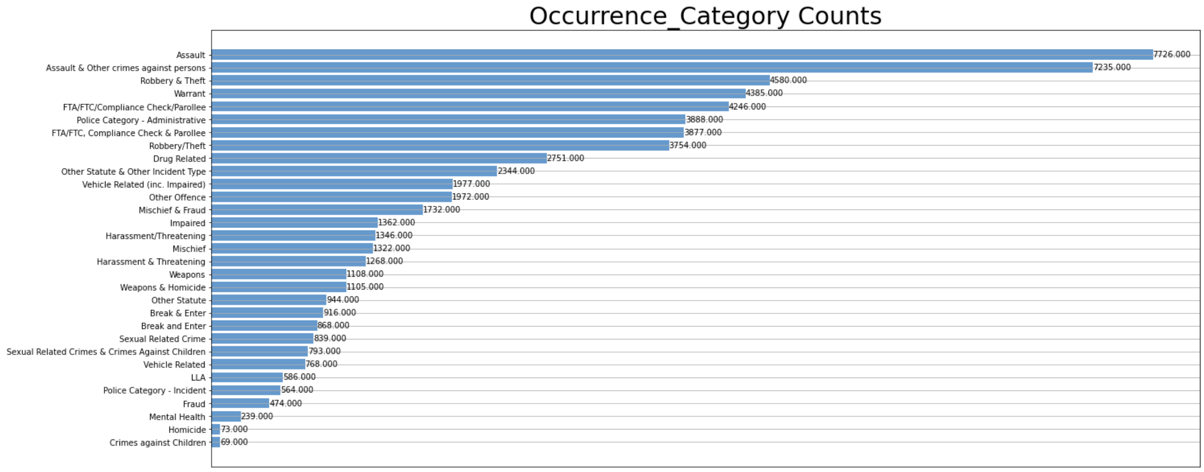
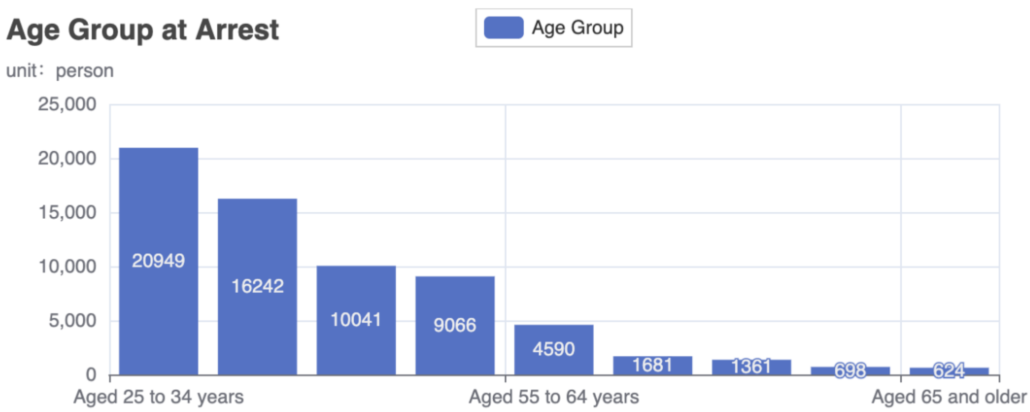


Figure 5 Occurrence Category Counts



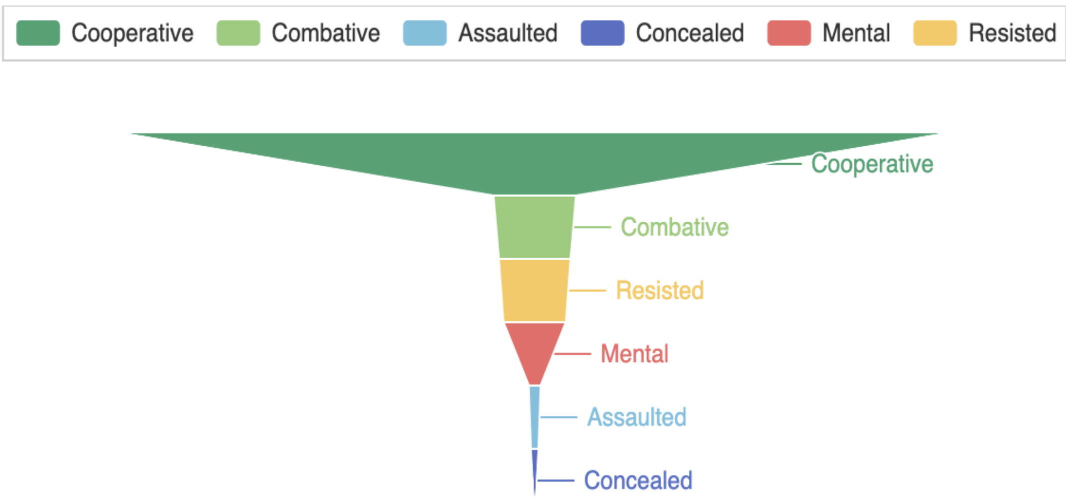
The next variable we study is the Age_group_at_arrest, which breaks the total observation based on different age levels. The following motivation graph (Figure 6) is a bar chart of the count of arrested or searched individuals by age group. Figure 6 shows that the age group of 25 to 34 years has the highest count, whereas the oldest age group (65 and older) has the lowest count.

Figure 6 Age Groups at Arrest



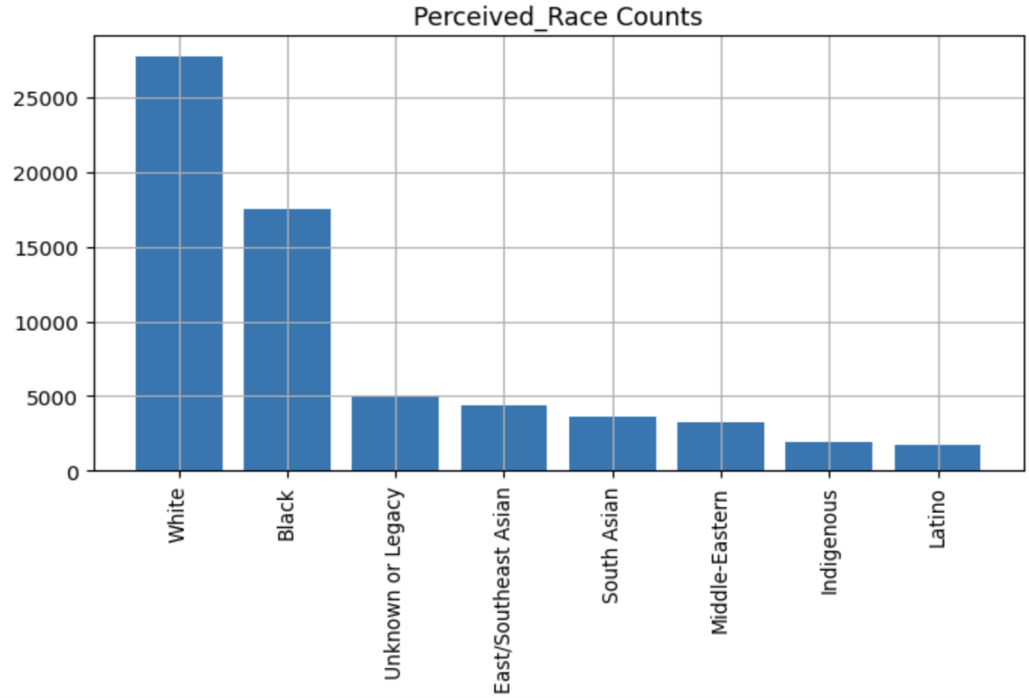
Moreover, we are interested in the proportional distribution of which combative actions happen most frequently. Therefore, we created a data frame to integrate all the variables of the actions at arrest and produced the following funnel graph to visualize the counts of each action happening in 2020 and 2021. Figure 7 shows that most of the sample population chose to cooperate with investigation and search, whereas only a small proportion would be triggered with combative actions.

Figure 7 Funnel Graph of Action at Arrest



Last, ‘Perceived_Race’ is a string variable with a list of races recorded by the police, including White, Black, East/Southeast Asian, Indigenous, Latino, Middle-Eastern, South Asian, and Unknown or Legacy. By observing Figure 8, we can see that the distribution of the counts of each perceived race is that white is the most commonly observed, whereas Latino is the least observed.

Figure 8 Counts of Perceived Race



T-test

Because Figure 6 shows only the distribution of count for the 9 Age Group at Arrest and the funnel distribution graph (Figure 7) for the 6 Actions at Arrest of the data frame, so we want to run Welch's t-test with categorical attributes in the dataset. Before running the tests, we checked that the following assumptions were fulfilled: (1) a nominal two-level explanatory variable; (2) a quantitative outcome variable; (3) a normality assumption; and (4) independence of the errors. As we ran Welch's t-test, equal variance among the residuals was not assumed. Therefore, we conducted the Welch T-test to analyze whether the specific means of the number of Actions at Arrest (outcome variables) differ between the Age Group of 25 to 34 years and 35 to 44 years (explanatory variable). The Hypothesis being tested below:

H0 (Null Hypothesis): The population means of two independent groups, people aged 25 to 34 years and 35 to 44 years, who acted differently (based on six different actions at arrest), are equal.
HA (Alternative Hypothesis): The population means of two independent groups, people aged 25 to 34 years and 35 to 44 years, who acted differently (based on six different actions at arrest), are not equal.

Table 1 T-test results based on Actions at Arrest and Aged 25 to 34 years and 35 to 44 years at Arrest in Arrest & Strip Searches dataset

Groups	Sample Size	Mean	SD	CI	DF
Aged_Group_at_Arrest (25-34)	12092	2015.33	3508.677111	(0.00, 898.00)	10
Aged_Group_at_Arrest (35-44)	9398	1566.33	2784.134671	(0.00, 898.00)	10

Table 2 T-test results of comparison of Action at Arrest frequency between Aged 25 to 34 years and 35 to 44 years at Arrest in Arrest & Strip Searches dataset

Test	Statistic	P-value	Conclusion
T-test on Aged_Group_at_Arrest (25-34 & 35 -44)	0.245545773	0.810999815	Fail to reject H0 as p-value is larger than 0.05

The results of Tables 1 & 2 indicate that the mean arrested people aged 25 to 34 with all actions at arrest size less than ten (10) (M=2015.3, 3508.68) is higher than those aged 35 to 44 years (M=1566.33, 2784.13). With alpha established at 0.05, there is no significant difference as the p-value (0.811) is more extensive than 0.05, 95% CI [0.00, 898.00]. Therefore, we fail to reject the null hypothesis that there is no difference in the means between arrested people aged 25 to 34 years (10) and 35 to 44 years acted differently (based on the six different actions ar arrest, 10).

Power Analysis

Figure 6, a bar chart illustrating the count of arrested or searched individuals by age group, indicates differences in the number of individuals across age groups. The highest count is found in the 25 to 34 age group, while the lowest is in the 65 and older age group. This finding suggests that the age group may influence the likelihood of being strip-searched. However, a power analysis is required to ascertain this relationship and quantify the effect size.

We can estimate the required sample size by conducting a power analysis to provide sufficient statistical power to detect an effect. This is important because an inadequately powered study may fail to identify a significant effect, resulting in a Type II error, also known as a false negative.

Moreover, it estimates the minimum sample size required to achieve the desired power level, optimizing the use of resources and minimizing the burden on participants. Strengthen the research design by quantifying the relationship between the age group at arrest and the likelihood of being strip-searched, thus allowing for more robust conclusions.

Method

We focus on data preparation in the first stage of our analysis approach. We select the relevant columns (‘Age_group_at arrest_’ and ‘SttripSearch’) from the original dataset, creating a new data frame named ‘df2’. This data frame will be used for the subsequent analysis.

Next, we move on to data preprocessing. In this step, we convert the categorical variables of ‘Age_group_at arrest_’ and ‘StripSearch’ in the dataset into numerical representations using the LabelEncoder from scikit-learn library. This transformation is necessary because many machine learning algorithms work with numerical data, and this step enables us to apply these techniques to our dataset. Additionally, we store the mapping of categorical labels to numerical values in the ‘res’ dictionary (Table 3) for future reference and showing below.

Table 3 Encoding Categorical Labels to Numerical Values

res	Mapping of Categorical labels to numerical values
Aged 17 years and under	0
Aged 17 years and younger	1
Aged 18 to 24 years	2
Aged 25 to 34 years	3
Aged 35 to 44 years	4
Aged 45 to 54 years	5

Aged 55 to 64 years	6
Aged 65 and older	7
Aged 65 years and older	8
nan	9

The third stage of our analysis approach is to estimate the effect size. For this purpose, we employed Cohen's d, a widely-used metric for determining the standardized difference between two means. We used two functions, 'pool_standard_deviation,' which calculates the pooled standard deviation of two samples, and 'Cohens_d,' which calculates the actual Cohen's value for independent samples. Therefore, we used these functions to compute the effect size of the age group at the time of arrest on the likelihood of being strip-searched.

The third stage of the analysis approach is to estimate the sample size needed to achieve a given statistical power, which in this case is 0.8. We will use the TTestlngPower().solve_power function from the statsmodels library. This approximation aids in identifying the smallest necessary sample size for discerning a statistically meaningful impact. Moreover, to account for the different proportions of observations in the age groups, we will calculate the ratio of one sample against the other. This ratio will be used as input in the solve_power function.

In the last phase of our analytical methodology, we generate a power curve graphic that illustrates the connection between the sample size and statistical potency across various effect magnitudes (0.2, 0.5, and 0.8). The plot_power function from the TTestlndPower class generates the power curve.

Results

Table 5 Cohen's D Effect Size for the Impact of Age Group on StripSearch

Cohen's D	Effect Size
Age group effect on StripSearch	0.120760061

We first calculate the effect size of the age group at the time of arrest on the likelihood of being strip-searched using Cohen's d metric. The obtained effect size was 0.128, indicating a slight standardized difference in the mean age group at arrest between strip-searched and those who were not.

Table 5 Needed sample Size and Actual Size for StripSearch and No-StripSearch Groups

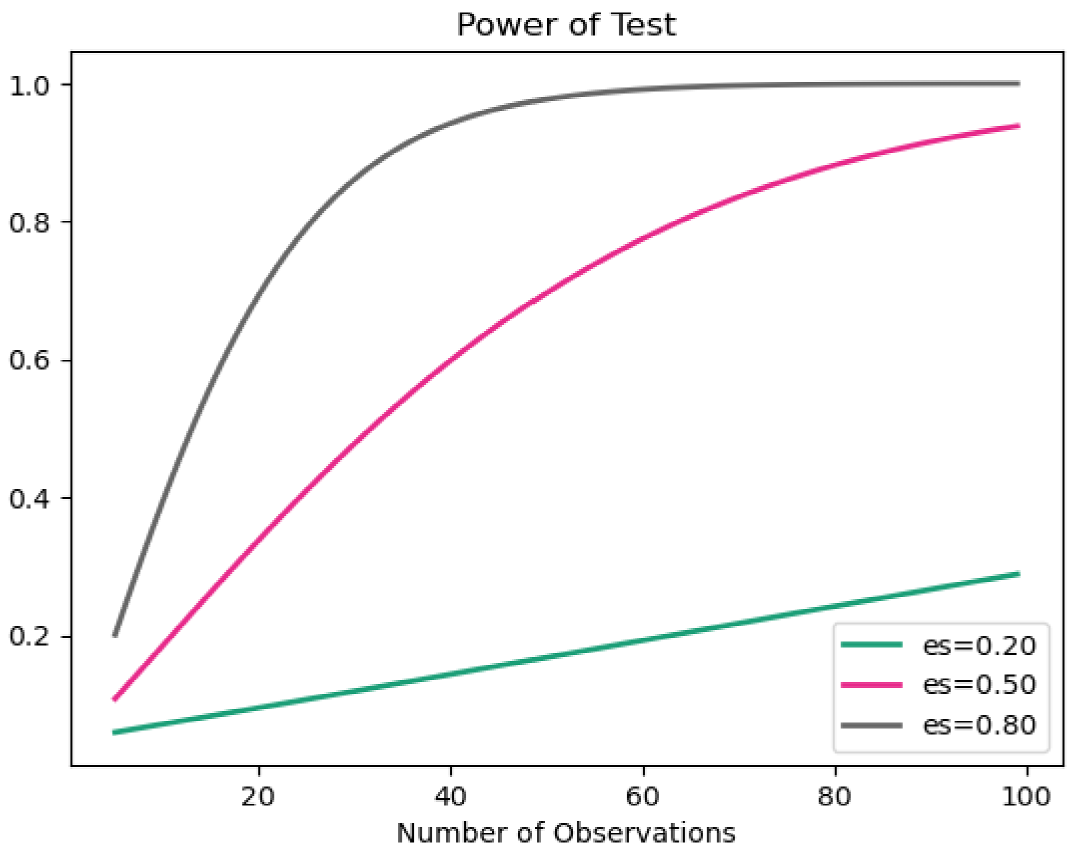
Group	Needed Sample Size	Actual Size
No-StripSearch (nobs1)	611.502	57475
StripSearch (nobs 2)	4505.327	7801

After obtaining the effect size, we estimate the required sample size by setting the desired statistical power at 80%(0.8) and the significance level at 5%(0005). The results indicate that a sample size of 611.502 is required for the group without strip-search(nobs1), while a sample size of 4505.327 is required for the group with strip-search (nobs2). It is crucial to note the relationship between the sample sizes of the groups, with nobs2 being considerably larger than nobs1. This proportion is consistent with the ratio observed in the original dataset.

Comparing these required sample sizes to the actual sample sizes available in the dataset, we see that the number of observations for the group without strip-search (No-StripSearch) is 57,475. For the group with strip-search (StripSearch), it is 7,801. The actual sample sizes for both groups considerably exceed the required sample sizes of 611.502 for the group without strip-search and 4505.327 for the group with strip-search.

Then we calculate the statistical power for a t-test using the TTestPower function from the statsmodels library. This calculation uses an effect size of 0.8, an alpha value of 0.05, and a sample size of 57,475 (the actual sample size for the group without strip-search). The calculated power is 1.000, which indicates that the test has a high probability of detecting a true effect of the specified size, given the large sample and effect sizes.

Figure 9 Power of Test



Upon examining the power curve plot of Figure 9, we can observe the following:

- With an increase in the sample size, the statistical power amplifies correspondingly for a specified magnitude of the effect. This means larger sample sizes provide a higher probability of detecting a true effect if it exists.
- For smaller effect sizes (e.g., 0.2), the required sample size to achieve a specific power level is considerably larger than those required for larger effect sizes (e.g., 0.5 or 0.8).
- Conversely, achieving a high power level can be accomplished with smaller sample sizes when the effect size is large.

ANCOVA

To investigate the research question, "How does the age group of the individual at the time of arrest influence the actions taken during the arrest while controlling for the occurrence category of the arrest and the individual's perceived race?" we conducted an exploratory data analysis using various visualizations. Figures 5 and 6 show the counts distribution for Occurrence_Category and Age_group_at_arrest, respectively. Figure 7 displays a funnel graph of the actions taken during the arrest. Additionally, we examined the distribution of the Perceived_Race variable.

Our exploratory data analysis revealed differences in the count distribution between age groups and actions taken during the arrest. To further investigate these differences, we ran Welch's t-test on the Actions at Arrest between age groups 25-34 and 35-44. The t-test results suggested no significant difference between these age groups regarding the actions taken during the arrest (p-value > 0.05). However, the results may be limited because the t-test only compared two age groups and did not account for other factors like Occurrence_Category and Perceived_Race.

To address these limitations and obtain a more accurate understanding of the relationship between the age group at the time of arrest and the actions taken during the arrest, we decided to perform an ANCOVA. By using ANCOVA, we can control for the occurrence category of the arrest and the individual's perceived race while examining the influence of age groups on the actions taken during the arrest. This method allows us to account for potential confounding factors and better address the research question.

Method

Initially, we prepare the dataset, which includes variables such as actions taken during the arrest, occurrence category, perceived race, and age group at the time of the arrest. We install the ‘penguin’ library for conducting the ANCOVA analysis. Next, we create a new data frame, ‘a2’, containing the actions taken during the arrest, derived by selecting the maximum value across relevant columns. We also extract the covariates, ‘Occurrence_Category’ and ‘Perceived_Race,’ into a separate DataFrame, ‘a3’. The between-group variables, ‘Age_group_at_ arrest_,’ is extracted into another DataFrame, ‘a4’.

We concatenate ‘a2’, ‘a3’, and ‘a4’ to form a new DataFrame, ‘df2’, which includes the actions taken during the arrest, occurrence category, perceived race, and age group at the time of the arrest. To preprocess the categorical variable in ‘df2’, we utilize the LabelEncoder from the ‘sklearn’ library, converting categorical values into numerical labels for the ‘actions,’ ‘Occurrence_Category,’ and ‘Perceived_Race’ columns. The resulting data frame, with numerical labels, represents the original categorical values.

Additionally, we store the mapping rules for each variable in the ‘relu1’ list for decoding the numerical labels back into their original categorical values later, shown below (Table

6&7&8). With the data prepared and transformed, we can perform the ANCOVA analysis to examine the influence of the age group at the time of arrest on actions taken during the arrest while controlling for the occurrence category of the arrest and the individual's perceived race. The transformed and preprocessed dataset, 'df2', enables us to investigate the relationships between variables and address the research question.

Table 6 Actions_at_Arrest_Mapping

Variable	Category	Mapping
Actions_at_arrest	Assaulted_o	0
	Combative_	1
	Concealed_i	2
	Cooperative	3
	Mental_inst	4
	Resisted_d	5

Table 7 Occurrence_Category Mapping

Variable	Category	Mapping
Occurrence_Category	Assault	0
	Assault & Other crimes against persons	1
	Break & Enter	2
	Break and Enter	3
	Crimes against Children	4
	Drug Related	5
	FTA/FTC, Compliance Check & Parollee	6
	FTA/FTC/Compliance Check/Parollee	7
	Fraud	8
	Harassment & Threatening	9
	Harassment/Threatening	10
	Homicide	11
	Impaired	12
	LLA	13
	Mental Health	14
	Mischief	15
	Mischief & Fraud	16

	Other Offence	17
	Other Statute	18
	Other Statute & Other Incident Type	19
	Police Category - Administrative	20
	Police Category - Incident	21
	Robbery & Theft	22
	Robbery/Theft	23
	Sexual Related Crime	24
	Sexual Related Crimes & Crimes Against Children	25
	Vehicle Related	26
	Vehicle Related (inc. Impaired)	27
	Warrant	28
	Weapons	29
	Weapons & Homicide	30
	nan	31

Table 8 Perceived_Race Mapping

Variable	Category	Mapping
Perceived_Race	Black	0
	East/Southeast Asian	1
	Indigenous	2
	Latino	3
	Middle-Eastern	4
	South Asian	5
	Unknown or Legacy	6
	White	7
	nan	8

Results

First, let us define the null and alternative hypothesis for each of the factors: age group at the time of arrest, occurrence category, and perceived race.

H0 (Null Hypothesis): There is no significant influence of the age group at the time of arrest on the actions taken during the arrest when controlling for the occurrence category of the arrest and the individual's perceived race.

HA (Alternative Hypothesis): There is a significant influence of the age group at the time of arrest on the actions taken during the arrest when controlling for the occurrence category of the arrest and the individual's perceived race.

Table 9 ANCOVA Results on Actions Taken During Arrest

Index	Source	SS	DF	F	p-unc	np2
0	Actions_at_arrest	4.040722	8	0.928888	4.91E-01	0.000114
1	Occurrence_Category	23.302556	1	42.854644	5.94E-11	0.000656
2	Perceived_Race	1.209254	1	2.223882	1.36E-01	0.000034
3	Residual	35475.316635	65241	NaN	NaN	NaN

Based on the ANCOVA results in Table 9, we can provide the following statistical interpretation for each variable:

- Age_group__at_arrest_: The uncorrected p-value (p-unc) is 0.4909, which is greater than 0.05. This means that we cannot reject the null hypothesis that the age group at the time of arrest does not significantly influence the actions taken during the arrest when controlling for the occurrence category of the arrest and the individual's perceived race.

- Occurrence_Category: The uncorrected p-value is 5.939747e-11, which is smaller than 0.05. This indicates a statistically significant relationship between the occurrence category of the arrest and the actions taken during the arrest when controlling for the age group at the time of arrest and the individual's perceived race.

- Perceived_Race: The uncorrected p-value is 0.1359, greater than 0.05. This means that we cannot reject the null hypothesis that the individual's perceived race does not significantly influence the actions taken during the arrest when controlling for the age group at the time of arrest and the occurrence category of the arrest.

Practical interpretation: From the results, we can conclude that the age group at the time of arrest and the individual's perceived race do not have a statistically significant relationship with the actions taken during the arrest when controlling for the occurrence category of the arrest. On the other hand, the occurrence category of the arrest significantly influences the actions taken during the arrest, even after controlling for the other factors.

Logistic Regression Analysis

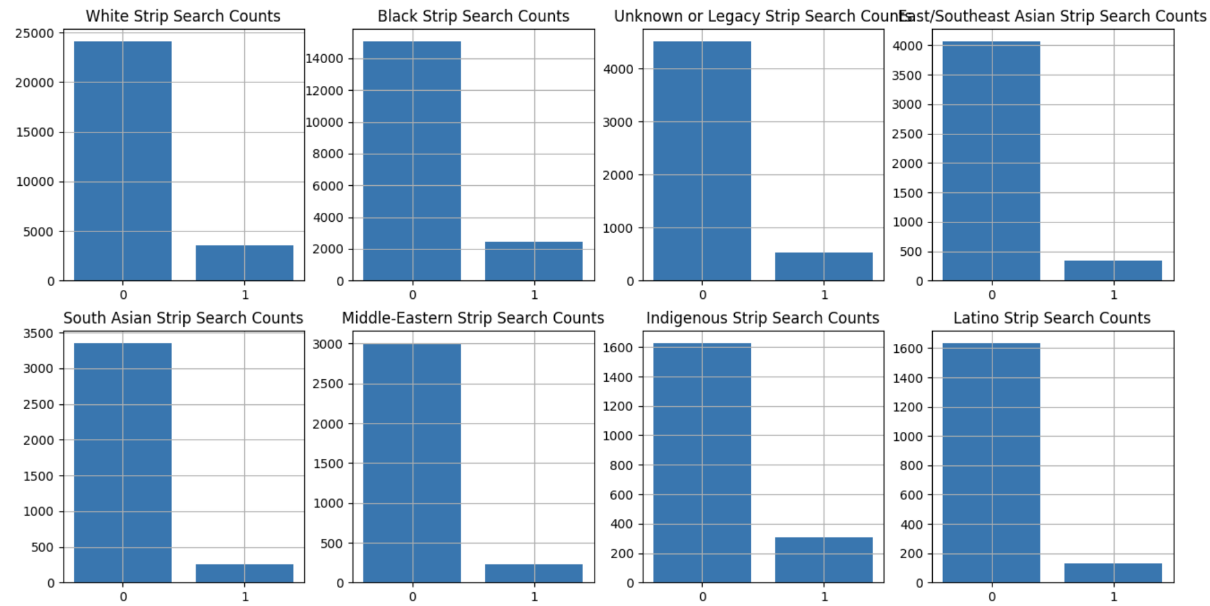
To study how perceived race affects the possibility of being strip-searched during an arrest, we conducted a logistic regression analysis based on the type of data and the research question. Utilizing the logistic regression approach, we could forecast the likelihood of a binary result by considering one or multiple independent factors. From a machine learning perspective, logistic regression is a supervised learning algorithm for binary classification tasks. Statistics is a simple yet powerful methodology that models the relationship between a set of input variables and a binary output (i.e., “yes” or “no”). With a logistic regression model, our goal is to find the best-fitting model that can predict the probability of the dependent variable by separating them into classes of 1 and 0 based on the input features.

Method

Since the logistic model is a machine learning model, and we started by identifying and preparing the data we wanted to include. Based on the research question, we chose ‘Perceived_Race’ as the independent variable and ‘StripSearch’ as the outcome variable and formed a data frame with these two variables as two columns.

We created a facet of each race's bar charts of strip searches. Figure 10 is another clear indication of the variables that we were looking into.

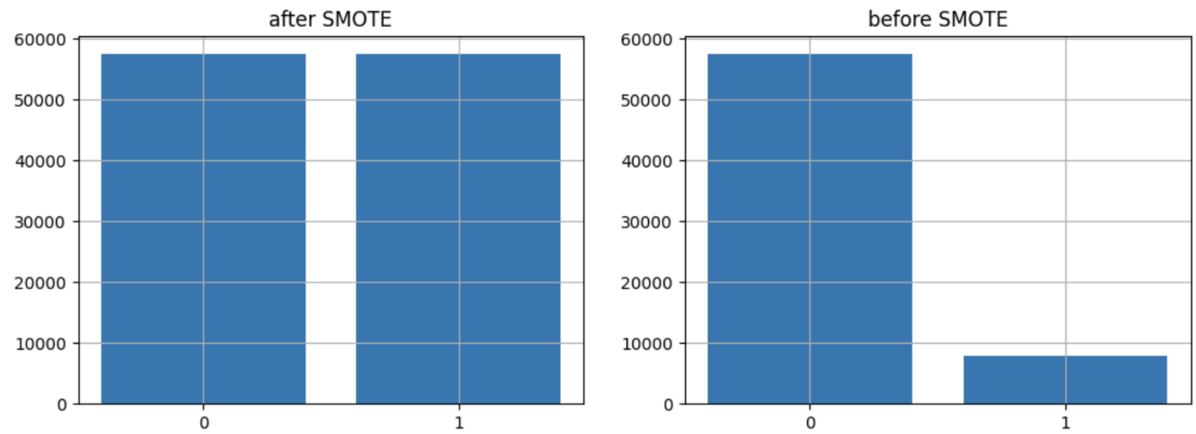
Figure 10 Facet of Bar Chart Strip Search by Race



Based on the feature of the independent variable, we split and created eight dummy variables for each race. For each variable, ‘1’ means this person is from the particular race the variable is representing, whereas ‘0’ means that this person is not from the race. In the case of logistic regression, the dependent variable, or the y variable, is also binary. For ‘StripSearch,’ ‘0’ represents the person not strip searched, whereas ‘1’ represents the person strip searched.

To train the data, we used the Synthetic Minority Over-sampling Technique (SMOTE) to balance the dataset since there is a significant imbalance between the total number of 0 and 1. Our goal is to address the class imbalance problem in the data frame, which could potentially negatively impact the model's performance. We used Python libraries to define the feature matrix 'X' by dropping the 'StripSearch' column from the data frame, defining the target variable 'y' as the 'StripSearch' column, and converting it to string data type. By applying the SMOTE technique, we created new variables, 'smote_X' and 'smote_y,' which are the feature matrix and target variable. Figure 11 shows the results by plotting the class distribution bar charts before and after SMOTE. As we can see from the left graph, the class distribution after SMOTE is much more balanced than the original distribution.

Figure 11 Class Distribution before/after SMOTE



After adjusting the imbalance issue, we split the adjusted dataset into training and testing groups, where 80 percent of the data points were used to train the model, and 20 percent were used to test the trained model's fitness. Based on this process, we used the *LogisticRegression model* from the *sklearn.linear_model* library to predict the results.

Results

After running the prediction model, we received an accuracy score of 0.5481 (54.81%), which indicates that the Logistic Regression model correctly classified approximately 54.81% of the instances in the test set. Based on the results of the logs of the odds of the model, the interpretation of the dependent variable:

- Given the race of the person, being black (Black = 1) would increase the chance of being strip-searched by the police by the predicted log of the odds at 1.1558 while holding other variables constant.
- Given the person's race, being East/Southeast Asian (Aisan = 1) would increase the chance of being strip-searched by the police by the predicted log of the odds at 0.6132 while holding other variables constant.
- Given the person's race, being Indigenous (Indigenous = 1) would increase the chance of being strip-searched by the police by the predicted log of the odds at 1.3252 while holding other variables constant.

- Given the person's race, being Latino (Latino = 1) would increase the chance of being strip-searched by the police by the predicted log of the odds at 0.5636 while holding other variables constant.
- Given the person's race, being Middle-Eastern (Middle-Eastern = 1) would increase the chance of being strip-searched by the police by the predicted log of the odds at 0.5505 while holding other variables constant.
- Given the person's race, being South Asian (South Asian = 1) would increase the chance of being strip-searched by the police by the predicted log of the odds at 0.5580 while holding other variables constant.
- Given the person's race, being Unknown or Legacy (Unknown or Legacy= 1) would increase the chance of being strip-searched by the police by the predicted log of the odds at 0.8460 while holding other variables constant.
- Given the person's race, being White (White= 1) would increase the chance of being strip-searched by the police by the predicted log of the odds at 1.0553 while holding other variables constant.

To further investigate the analysis, we visualized the feature importance of the Logistic Regression model by plotting the coefficients of each race in the model. Races with larger positive coefficients, such as Indigenous and Black, have a higher positive impact on the predictor. On the other hand, races with larger negative coefficients, such as East/Southeast Asian, Latino, Middle-Eastern, and South Asian, have a higher negative impact. Lastly, races with coefficients close to zero have a lower impact on the model's prediction.

Table 10 Multinomial Logistic Regression Coefficients and Odds Ratios for Perceived_Race

Variables	Coefficient	Odd Ratio
Black	0.1448	1.1558
East/Southeast Asian	-0.489	0.6132
Indigenous	0.2815	1.3252
Latino	-0.5733	0.5636
Middle-Eastern	-0.5969	0.5505
South Asian	-0.5833	0.558

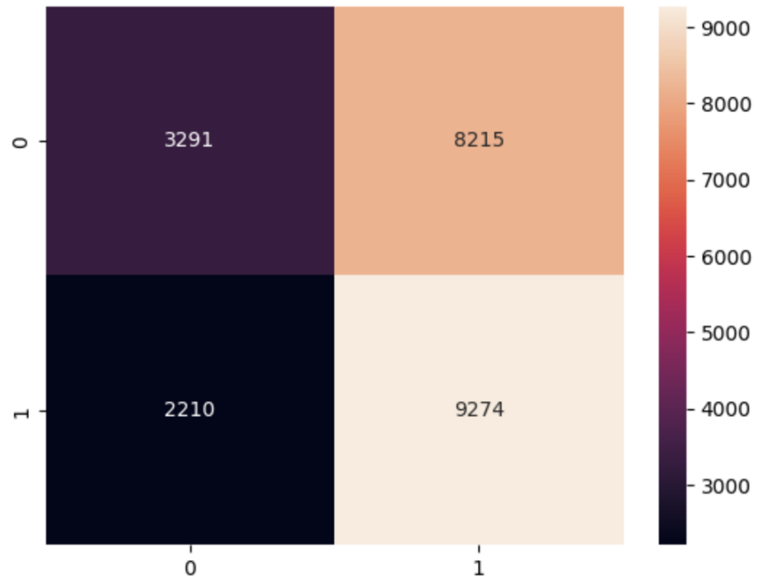
Lastly, we ran a confusion matrix for the prediction to testify the accuracy of the results.The heatmap map illustrates the quantity of accurate positive results (TP), correct negative outcomes (TN), incorrect positive detections (FP), and erroneous negative instances (FN) associated with the algorithm. In this case, the x-axis represents the predicted labels, while the y-axis represents the true labels. Given the confusion matrix data, we calculated the following scores for the model's performance:

- Accuracy: 54.65%

- Precision: 28.60%

The accuracy of 54.65% means that the model correctly predicted 54.65% of the instances, while the precision of the model is 28.60%. The interpretation and discussion of the result will be included in the discussion section.

Figure 12 Confusion Matrix



Discussion

Power Analysis

Our power analysis focused on the relationship between the age group at the time of arrest and the likelihood of being strip-searched. We first calculated the effect size (Cohen's D) for the age group's effect on strip-search is 0.1208, indicating a small effect according to common interpretation guidelines. We estimated the required sample sizes for both groups, with and without strip-search, to achieve a statistical power of 80% given the calculated effect size. The results demonstrated that a sample size of 611.502 for the group without strip-search and 4505.327 for the group with strip-search would be needed. Interestingly, the sample size for the StripSearch group was notably larger than the NO-StripSearch group, reflecting the ratio present in the original dataset.

Upon comparing the actual sample sizes in the dataset (57,475 for No-StripSearch and 7801 for StripSearch) with the required sample sizes, it became evident that both groups had sufficient sample sizes to detect the small effect of age group on the likelihood of being strip-searched.

The power analysis also included calculating the statistical power for a t-test using an effect size of 0.8, an alpha value of 0.05, and a sample size of 57,475 for the NO-StripSearch group. Given the large sample and effect size, the resulting power of 1.000 indicates a high probability of detecting a true effect.

The power curve visualization, generated with the plot_power function from the TTestIndPower class, illustrated the relationship between sample size, effect size, and statistical power. The power curve demonstrated that a larger sample size increases the probability of detecting a true effect, that smaller effect sizes require large sample sizes to achieve a specific power level, and that high levels of power can be attained with relatively smaller sample sizes when the effect size is large.

ANCOVA Analysis

The results from ANCOVA analysis indicate no significant relationship between the age group at the time of arrest and the actions taken during the arrest when controlling for the occurrence category of the arrest and the individual's perceived race. This finding suggests that an individual's age group at the time of arrest may not be a significant predictor of the actions taken by the police during the arrest.

Similarly, the individual's perceived race does not show a significant relationship with the actions taken during the arrest when controlling for the age group at the time of arrest and the occurrence category of the arrest. This result is an interesting insight, as it suggests that, at least within the context of this dataset, the individual's perceived race is not significantly related to the actions taken during the arrest.

In contrast, the occurrence category of the arrest exhibits a significant relationship with the actions taken during the arrest when controlling for the age group at the time of arrest and the individual's perceived race. This implies that the nature of the arrest (i.e., the type of offense) significantly influences the actions taken by the police during the arrest, regardless of the age or perceived race of the individual.

Logistic Regression Analysis

Based on the logistic regression results, we learned that the largest positive coefficient is 0.2518, with an odds ratio of 1.3252 for being Indigenous. In contrast, the largest negative coefficient is -0.1672, with an odds ratio of 0.8460. However, based on the preliminary research, the result must provide a more accurate prediction.

First, we are given an accuracy score of 54.65%, which is a low score, which means that the model may need to be more reliable. The calculation of an accuracy score measures the proportion of correct predictions over the total number of observations. Though we cleaned and adjusted the imbalanced dataset, the inconsistency in raw data still exists in the dataset might negatively impact the final results. The precision rate indicates that among the instances predicted as positive, only 28.60 percent were positive. This low precision shows that the classifier produces relatively high false positives or type 1 errors.

Improve the current model, and there are several methods we can further implement to improve accuracy. First, we could include more relevant variables in the model. As for now, the only variables in the model are dummy variables representing each race. By including other variables, such as numeric variables, we may derive better results by training the data and the model. Second, we can optimize the model's hyperparameters using techniques such as random search. This could help find the best combination of hyperparameters for the model instead of just using the default parameters. Last but not least, it would be better to run cross-validation on the model. Applying cross-validation to estimate the model's performance on unseen data could help us assess the model's generalizability and identify areas for improvement.

Conclusion

Based on the data introduction and literature, it is evident that strip searches are subject to debate regarding their use in police inspection. The study examined how an effect size of sampling, different Age groups, and different Perceived Races impact the potential result of getting stipe searched. Three research questions were formulated to achieve the analysis purpose. Preliminary research and an EDA were conducted to investigate data distribution and identify solutions to avoid bias before formal analysis.

The power analysis results revealed a small effect size of 0.128 for the influence of age group at the time of arrest on the likelihood of being strip-searched. While this effect size suggests a minor relationship between these factors, the analysis provided valuable insights into the importance of considering statistical power, effect sizes, and sample sizes in research design. By ensuring that our study had sufficient statistical power, we minimized the likelihood of committing a Type II error and increased the reliability of our findings. Future research may build upon these findings by further investigating the relationship between the age group at arrest and the likelihood of being strip-searched, exploring other potential factors that may influence this relationship, and utilizing more advanced statistical techniques to examine the complex interplay between various demographic and situational factors.

The ANCOVA analysis provides valuable insights into the factors influencing the actions taken during an arrest. The results suggest that the age group at the time and the individual's perceived race do not significantly affect the actions taken during the arrest when controlling for the occurrence category of the arrest. On the other hand, the occurrence category of the arrest significantly influences the actions taken during the arrest. The findings from ANCOVA analysis may be useful for informing policing strategies and interventions to improve the interactions between law enforcement officers and individuals during arrests. Policymakers and law enforcement agencies should consider the occurrence category of the arrest as a significant factor that influences the actions taken during the arrest process.

Additionally, the current analysis is limited by the dataset used, which may not represent all arrests and their outcomes. Further research should try replicating these findings using larger and more diverse samples to understand better the relationship between age, perceived race, occurrence category, and actions taken during an arrest.

We also conducted a multivariate logistic regression to examine whether being a particular race would increase the chances of being strip searched. The results showed that Indigenous people are most likely to get strip searched. However, due to the constraints in the data selection of independent variables, the imbalance in the dataset may cause the lack of generality in the trained model to perform prediction on new data. We identified three ways to further improve the model by adding more variables, optimizing the hyperparameters of the model, and conducting cross-validation on the model.

Our study explored the influence of age group, perceived race, and occurrence category on strip-search likelihood during an arrest. The power analysis revealed a small effect size for the age group, emphasizing the importance of sufficient statistical power. The ANCOVA analysis showed that the occurrence category significantly affects arrest actions, while age and perceived race do not. Policymakers and law enforcement agencies should consider these findings when improving arrest interactions. The multivariate logistic regression found that Indigenous people had the highest strip-search likelihood, but further research with larger samples is needed to confirm this. Future studies should also enhance the model's generality through additional variables, hyperparameter optimization, and cross-validation.

References

Arrests and strip searches (RBDC-arr-TBL-001). Toronto Police Service Public Safety Data Portal. (2022, November 10). Retrieved February 24, 2023, from https://data.torontopolice.on.ca/datasets/TorontoPS::arrests-and-strip-searches-rbdc-arr-tbl-001/about

Asa Series on How Race and Ethnicity Matter. American Sociological Association - Department of Research and Development. (2007, September). Retrieved February 25, 2023, from https://www.asanet.org/wp-content/uploads/savvy/images/press/docs/pdf/ASARaceCrime.pdf

Global News. (2023, February 24). *Toronto police statistics show disproportionate use of force on black people*. Global News. Retrieved February 28, 2023, from https://globalnews.ca/news/8921745/toronto-police-release-race-based-data-force-strip-searches/

McNeilly, G. (2019, March). Breaking the golden rule - office of the independent police review director. Retrieved February 25, 2023, from https://www.oiprd.on.ca/wp-content/uploads/OIPRD_Breaking-the-Golden-Rule_Report_Accessible.pdf

Tanovich, D. (2011, February). Bonds: Gendered and Racialized Violence, Strip Searches, Sexual Assault and Abuse of Prosecutorial Power. Retrieved February 25, 2023, from https://www.uwindsor.ca/law/tanovich/

Toronto Police Service. (2022, June 15). Race-based data collection. ArcGIS StoryMaps: Race-Based Data Collection. Retrieved February 28, 2023, from https://storymaps.arcgis.com/stories/d648e2827cb74ea6bda9c96bd22b5c08

Ulmer, J., & Steffensmeier, D. (Eds.) (2014). The age and crime relationship: Social variation, social explanations. SAGE Publications Ltd, https://dx.doi.org/10.4135/9781483349114

Written on Behalf of Barrison Law. (2021, April 6). Strip searches in Ontario are occurring too often. Barrison Law. Retrieved February 24, 2023, from https://criminallawoshawa.com/strip-searches-in-ontario-are-occurring-too-often/