INF 2178H S LEC0101 20231
Experimental Design for Data Science

Final Project

The Analysis of
Arrests and Strip Searches
Group 8

Yuetong Jiang 1004339343
Jingyu Jia 1009725969

APR 15th 2023

Table of Content

Introduction **3**
- Literature Review 3
- Dataset Description 5
- Research Objective and Questions 6

EDA **7**

Method **16**

Results and Findings **16**
- Power Analysis 17
- ANCOVA 18
- Logistic Regression 19

Discussion **27**
- Limitation 28

Conclusion **29**

Appendix **30**

Introduction

Race-Based Data Collection, Analysis and Public Reporting Policy1 (Board Policy) dated September 2019 was approved by the Toronto Police Services Board on On September 19, 2019 to identify, monitor and address systemic disparities in policing. The Toronto Police Service Board had alleged to the public that they'll always do all the best to maintain fairness in all kinds of policing practices.

However, disparities are still the issues that people complain the most, especially related to police strip research incidents. Therefore, our team decides to investigate the potential relationships between various demographic factors and policing practices, with a particular focus on the degree of aggressiveness exhibited by criminals and the use of strip searches.

Based on the results obtained from T-tests, Tukey's Test, ANOVA tests, and Linear Regression, we conclude that age group and gender do account for disparities in terms of policing practices especially when police did strip research.

Literature Review

Phan, M. B., Panaitescu, M., & Rebelo, N. (2022, February). *Race & Identity Based Data Collection Strategy*. Online Reporting- Toronto Police Service. Retrieved February 26, 2023, from https://www.tps.ca/services/online-reporting/

On September 19, 2019, the Toronto Police Service (Board) approved the Race-Based Data Collection, Analysis and Public Reporting Policy dated September 2019 to identify, monitor and address systemic racial disparities in policing.

Current situation at that time :

Police Search of Persons in Arrests Service procedure 01-02 Search of Persons governs and outlines possible risk factors for the search of persons, but the decision as to what type of search is appropriate should be accessed on a case-by-case basis. There are four levels of searches: protective, frisk, strip search, and body cavity and all searches must be done sequentially, and there must be reasonable justification for higher levels of searches.

The definition of strip search used by the Service is:

A thorough search of a person's clothing and non-physical search of the body that will often require removal or rearrangement of some, or all, of the person's clothing to permit a visual inspection of a person's private areas: namely the genitals, buttocks, breasts or chest, body cavity, and/or undergarments; the mouth was excluded from this definition despite being a bodily cavity.

(*R. vs. Golden 2001 SCC 83*)

In October 2020, the Service revised the Procedure 01-02, *Search of Persons*. The changes address office accountability, training, and data management. Some of these changes include:

1. All protective and frisk searches are captured on audio and video, wherever possible, to allow for transparency and accountability.
2. All strip searches must be authorized by a supervisor and documented.
3. Update of the information captured during a strip search.
4. All strip searches must be audited at a divisional and senior management level.
5. All officers must complete a robust training module of search of persons including a review of case law.

Then, author Dr. Mai B. Phan, Mihaela Dinca-Panaitescu and Nicole Rebelo conducted *Understanding Strip Searches in 2020 Methodological Report* in October 2022 to analyze the outcome of a change in policy of strip search. The following are the major findings.

To summarize their findings:

1. Strip search rates varied throughout 202 and were significantly reduced following changes in Search policy and procedures in October 2020.
2. Indigenous, Black and White people were over-represented in strip searches relative to their proportion in arrests. After policy changes, there was no longer any over-representation of Indigenous people in strip searches.
3. There were differences by race after accounting for repeat arrests, and arrests related to drug and weapons offenses.
4. There were differences in strip search rates across the city where arrests took place.

Dataset Description

The dataset used for analyses comes from the Records Management System (R.M.S) that is used to record situations about arrests and strip searches. In normal procedure, people must first be arrested for a violation of law and then may be subject to a strip search. They also must be booked into custody at a police station if they are subject to a strip search. Occurrence records are sources of information about incidents including situations and individuals who were arrested.

The following table is the description of the dataset:

Dataset Name	**Type**	**Description**
Arrest_Year	Number	Year when a person is arrested
Arrest_Month	Number	Month period when a person is arrested
EventID	Number	Identification number of an event
ArrestID	Number	Identification number of an arrest
PersonID	Number	Identification number of a person
Perceived_Race	Text	Types of races of an arrested person
Sex	Text	Type of sex an arrested person
Age_group__at_arrest	Text	Age group of an arrested person
Youth_at_arrest__under_18_years	Text	Whether the age of a arrested person is a youth (age<18)
ArrestLocDiv	Text	Aggregated to the Division level and refers to where the arrest took place within Division boundaries. Marked xx when arrest took place outside City of Toronto
StripSearch	Number	If strip search is conducted, 0 means no, 1 means yes.
Booked	Number	If custody is booked, 0 means no, 1 means yes.
Occurrence_Category	Text	The types of occurrence category
Actions_at_arrest__Concealed_i	Number	If concealed items happened at arrest, 0 means no, 1 means yes.

Actions_at_arrest__Combative__	Number	If combative, violent or spitter/biter happened at arrest, 0 means no, 1 means yes.
Actions_at_arrest__Resisted__d	Number	If resisted, defensive or escape risk happened at arrest , 0 means no, 1 means yes.
Actions_at_arrest__Mental_inst	Number	If Mental instability or possibly suicidal, 0 means no, 1 means yes.
Actions_at_arrest__Assaulted_o	Number	If assaulted officer happened at arrest,0 means no, 1 means yes.
Actions_at_arrest__Cooperative	Number	If cooptative happened at arrest, 0 means no, 1 means yes.
SearchReason_CauseInjury	Number	If cause injury is a search reason, 0 means no, 1 means yes.
SearchReason_AssistEscape	Number	If assist escape is a search reason,0 means no, 1 means yes.
SearchReason_PossessWeapons	Number	If possess weapons is a search reason, 0 means no, 1 means yes.
SearchReason_PossessEvidence	Number	If possess evidence is a search reason, 0 means no, 1 means yes.
ItemsFound	Number	If item is found,0 means no, 1 means yes.
ObjectId	Number	Identification number of a found item.

Table 1: Dataset Description

Research Objective and Questions

The research objective of this study is to investigate the potential relationships between various demographic factors and policing practices, with a particular focus on the potential demographic factors which could influence the probability of being strip searched. By exploring these factors and relationships, we aim to better understand the potential discriminatory practices that may exist in law enforcement and to suggest areas for improvement.

RQ1: Will the probability of strip search be influenced by different gender groups?

RQ2: Will the probability of strip search be influenced by people's gender, age, race, and their actions when arrested?

The first research question seeks to understand whether the probability of being strip searched are influenced by one's age when controlling their actions at arrest. For question 2, we aim to investigate whether there is a correlation between an individual's gender, age, race, and behavior during arrest, and the likelihood of being subjected to a strip search. This is an important topic to explore because strip searches can be traumatic experiences that should not be used as a means to discriminate based on personal prejudices. By examining the existence of discrimination, we can work towards more equitable and just policing practices that promote fairness and better serve all members of society.

EDA

1. Strip Search & Sex

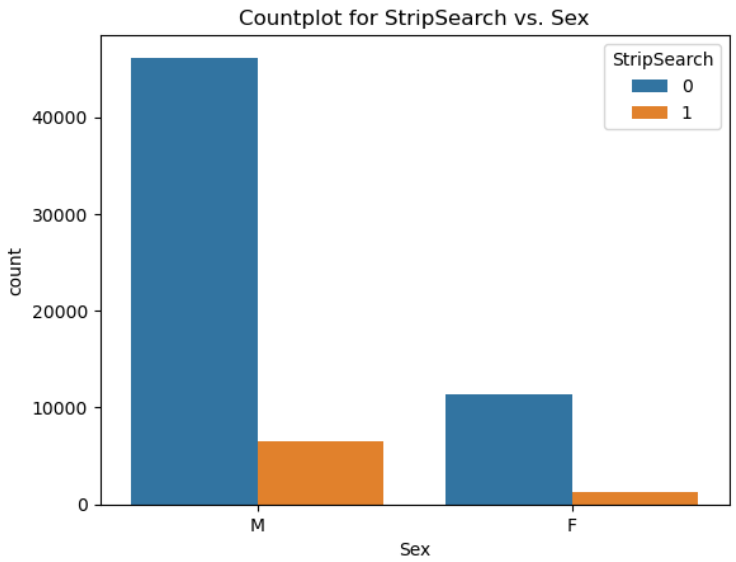

Figure 1: Countplot for StripSearch vs. Sex

As we can see in Figure 1, the database has a higher percentage of male criminals and also more male individual cases of being searched.

T-test:

H0: the mean of strip search for male and female are not different

H1: the mean of strip search for male and female are different

Check Assumptions:

a. Independence: Yes, each measurement is independent on each observation

b. Normal distributed: Not really, this is one of our model limitation
c. Equal variance: Yes. We use Welch's T-test which doesn't require equal variance assumption to be met.

With a pre-established alpha level of 0.05, we can conclude that there is a statistically significant difference between the probability of strip search for males and females. The test statistic is -7.25, and the p-value (<0.001) is less than the alpha level. Therefore, we can reject the null hypothesis.

Chi-Squared Independence Test:

H0: Sex and Strip Search are independent (no relationship)

Ha: there is relationship between Sex and Strip Search

Check Assumptions:

a. random sample: Yes, all data are randomly selected and is representative of the population
b. Categorical variables: Yes, both strip search and sex are categorical variables
c. Sufficient sample size: Yes, based on our contingency table below, all entries are greater than 10.

	Not Strip Searched	Strip Searched
Female	11334	1283
Male	46132	6518

Table 2: Contingency Table for Sex & Strip Search

With a pre-established alpha level of 0.05, we can conclude that there is a statistically significant relationship between sex and strip search. The test statistic is 47.07, and the p-value (<0.001) is less than the alpha level. Therefore, we can reject the null hypothesis.

2. Strip Search & Age

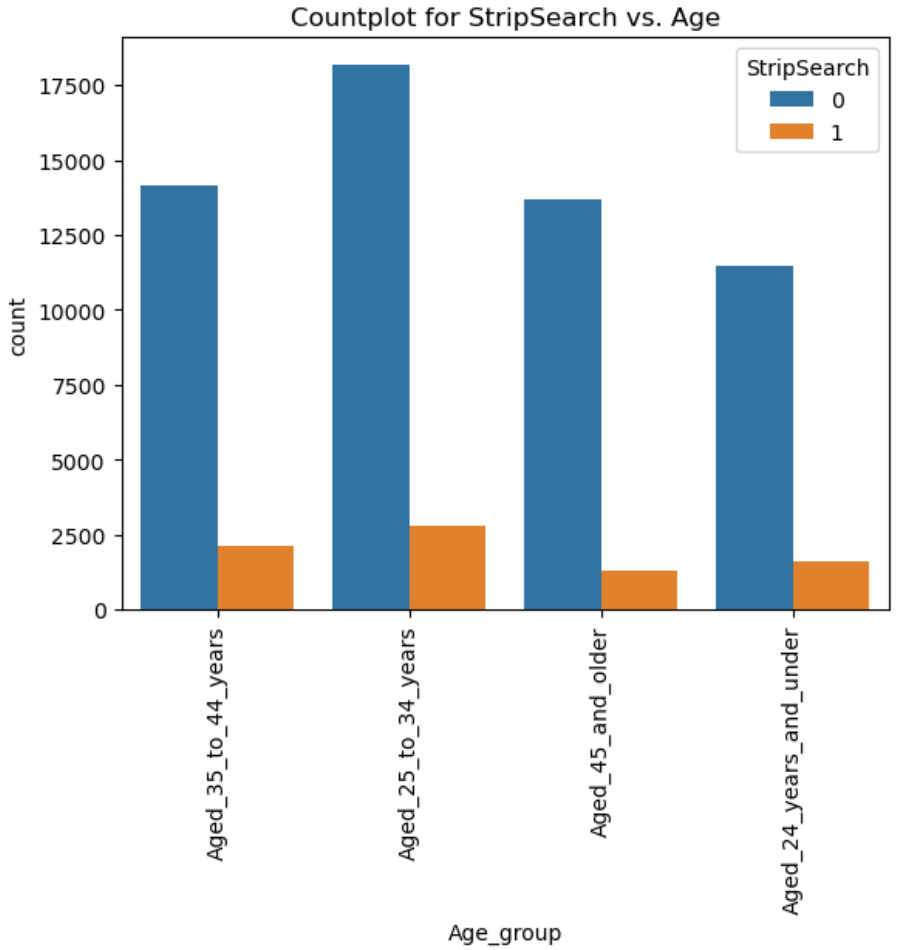

Figure 2: Countplot for StripSearch vs. Age

From Figure 2, we can find there exists some difference in strip search amount between age groups. With the same number of people not searched, more criminals under the age of 24 were searched than criminals aged 45 or older.

Tukey's test:

Check Assumptions:

a. Independence: Yes, each measurement is independent on each observation
b. Normal distributed: not really, this is one of our model limitation
c. Equal variance: Yes based on our summary statistics

We have 6 hypothesis tests, for each of them:

H0: the pair of age groups have the same mean strip search probability
H1: the mean of strip search for this pair of age groups are different

The age category has 4 different levels: 'Aged 35 to 44 years', 'Aged 25 to 34 years', 'Aged 45 and older', 'Aged 24 years and under'. The Tukey's HSD test compared each pair of them to check the difference between means. Therefore, there are a total six hypothesis tests, where the null hypotheses are that the mean differences between each pair of groups equal to 0, $\mu_2 - \mu_1 = 0$ and alternative hypotheses are differences not equal to 0, $\mu_2 - \mu_1 \neq 0$.

Multiple Comparison of Means - Tukey HSD, FWER=0.05

group1	group2	meandiff	p-val	lower	upper
Aged 24 years and under	Aged 45 and older	-0.0375	<0.001	-0.0475	-0.0276
Aged 25 to 34 years	Aged 45 and older	-0.0453	<0.001	-0.0542	-0.0364
Aged 35 to 44 years	Aged 45 and older	-0.043	<0.001	-0.0516	-0.0327

Table 3: Tukey's HSD Test Result: Age Group and Strip Search

We observe that three out of six comparisons in Table 3 are statistically significant, with p-values less than 0.05. Specifically, there are significant differences in the probability of strip search between the 'Age>=45' group and all three other age groups. Hence, we can conclude that there is a statistically significant difference in the probability of strip search between age groups.

Chi-Squared Independence Test:

Check Assumptions:

a. random sample: Yes, all data are randomly selected and is representative of the population
b. Categorical variables: Yes, both strip search and sex are categorical variables
c. Sufficient sample size: Yes, based on our contingency table below, all entries are greater than 10.

	Not Strip Searched	Strip Searched
Aged <= 24	11452	1629
Aged 25 - 34	18174	2771
Aged 35 - 44	14143	2098
Aged >= 45	13673	1303

Table 4: Contingency Table for Age & Strip Search

With a pre-established alpha level of 0.05, we can conclude that there is a statistically significant relationship between age and strip search. The test statistic is 200.4, and the p-value (<0.001) is less than the alpha level. Therefore, we can reject the null hypothesis.

3. Strip Search & Race

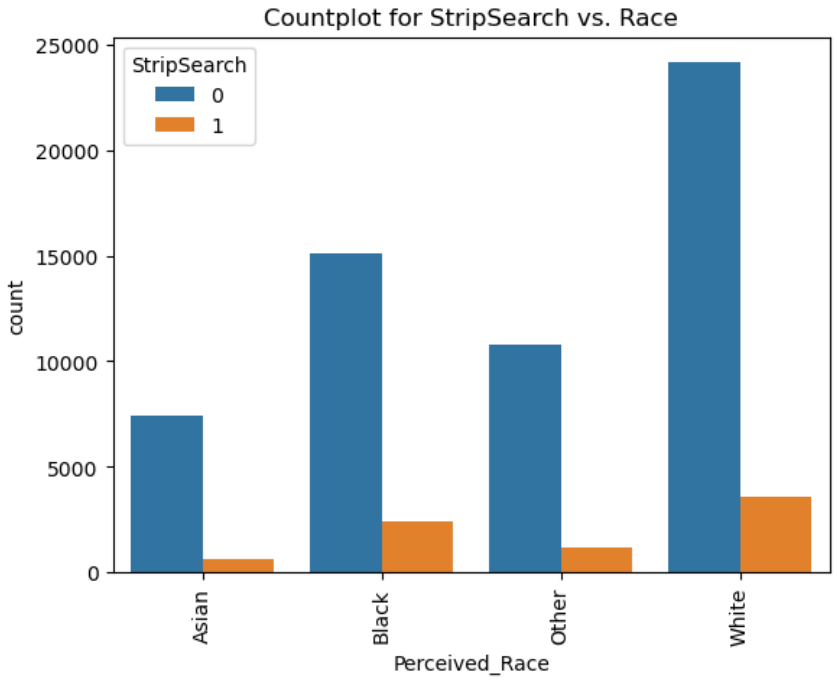

Figure 3: Countplot for StripSearch vs. Race

From Figure 3, we can find there exists some difference in strip search amount between Race groups. With a similar number of people being searched, more white criminals were not searched than Black criminals.

Tukey's test:

Check Assumptions:

a. Independence: Yes, each measurement is independent on each observation
b. Normal distributed: not really, this is one of our model limitation
c. Equal variance: Yes based on our summary statistics

We have 6 hypothesis tests, for each of them:

H0: the pair of Race groups have the same mean strip search probability

H1: the mean of strip search for this pair of Race groups are different

The race category has 4 different levels:'White', 'Other', 'Black', 'Asian'. The tukey's HSD test compared each pair of them to check the difference between means. Therefore, there are a total six hypothesis tests, where the null hypotheses are that the mean differences between each pair of groups equal to 0, $\mu_2 - \mu_1 = 0$, and alternative hypotheses are the mean differences not equal to 0, $\mu_2 - \mu_1 \neq 0$.

Multiple Comparison of Means - Tukey HSD, FWER=0.05

group1	group2	meandiff	p-val	lower	upper
Asian	Black	0.0644	<0.001	0.0532	0.0756
Asian	Other	0.0258	<0.001	0.0138	0.0378
Asian	White	0.0542	<0.001	0.0436	0.0647
Black	Other	-0.0386	<0.001	-0.0485	-0.0288
Black	White	-0.0102	0.0058	-0.0183	-0.0022
Other	White	0.0284	<0.001	0.0193	0.0375

Table 5: Tukey's HSD Test Result: Race and Strip Search

All six tests in Table 5 are statistically significant, with p-values less than 0.05. Thus, we can conclude that there is a statistically significant difference in the probability of strip search between race groups.

Chi-Squared Independence Test:

Check Assumptions:

a. Random sample: Yes, all data are randomly selected and is representative of the population
b. Categorical variables: Yes, both strip search and sex are categorical variables
c. Sufficient sample size: Yes, based on our contingency table below, all entries are greater than 10.

	Not Strip Searched	Strip Searched
Asian	7427	598
Black	15084	2434
Other	10789	1203
White	24142	3566

Table 6: Contingency Table for Race & Strip Search

With alpha established at 0.05, we can conclude that there is a statistically significant relationship between race and strip search. The test statistic is 281.35, and the p-value (<0.001) is less than the alpha level. Therefore, we can reject the null hypothesis.

4. Sex, Age, Race & Strip Search

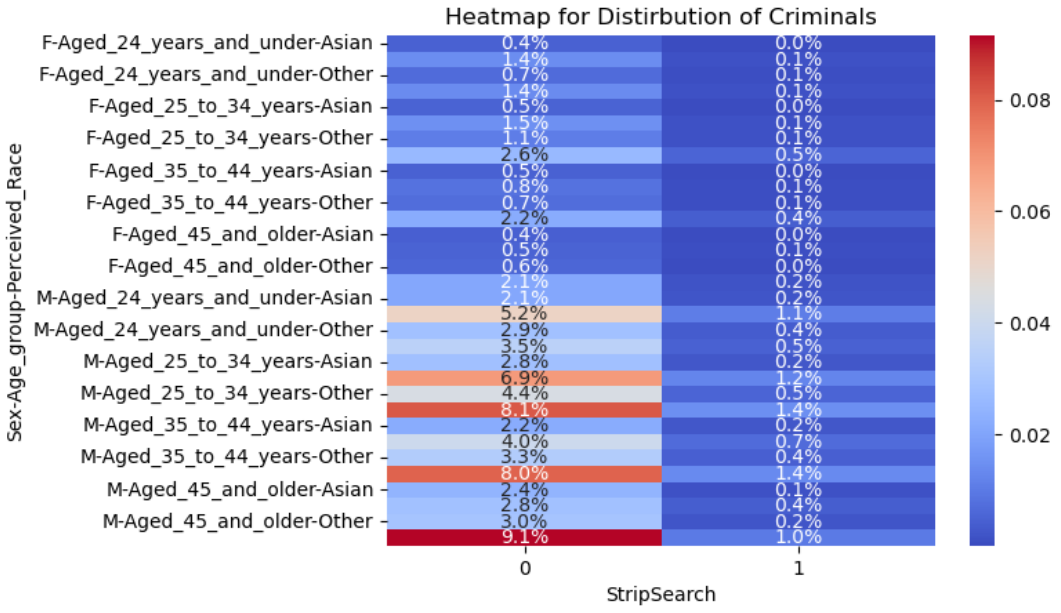

Figure 4: Heatmap for Distribution of Criminals

Figure 4 shows the proportion of people not being strip searched and being strip searched for each possible group. As we can find, most of the criminals were not strip searched, where the criminals who are Male, Other race, aged 45 and older accounted for the largest proportion of all criminals.

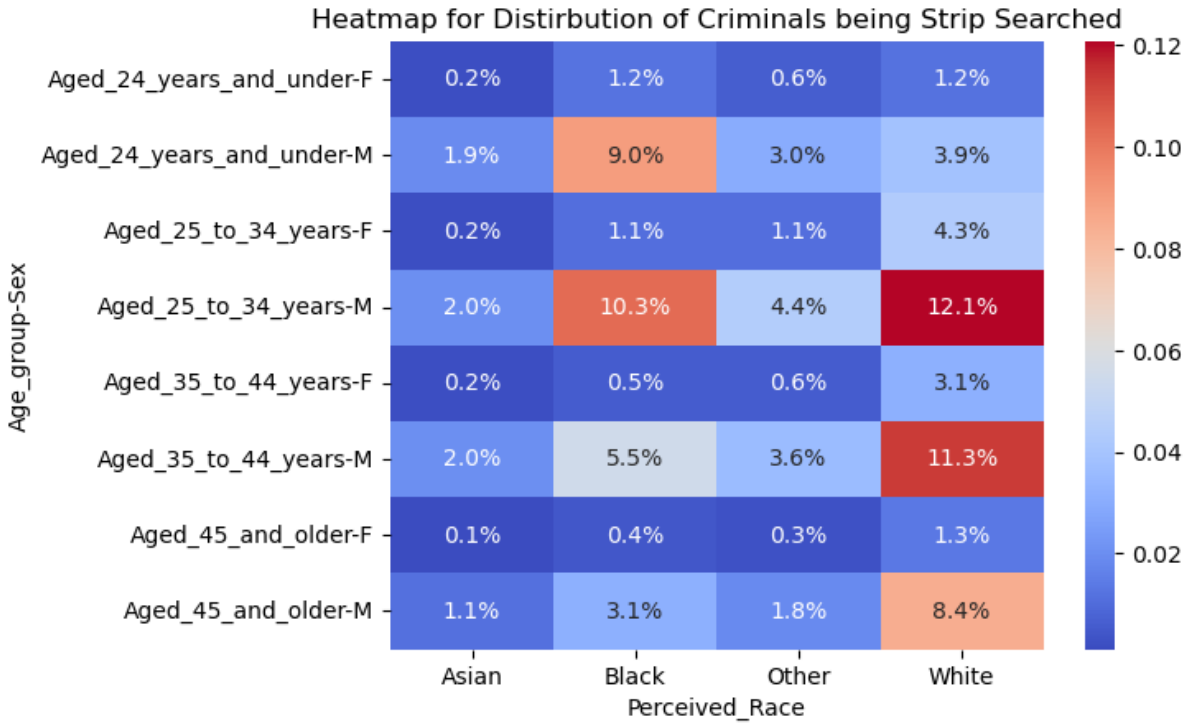

Figure 5: Heatmap for Distribution of Criminals Being Strip Searched

As shown in Figure 5, for those who are strip searched, criminals who are white, male, aged 25 to 34, aged 35 to 44, and aged 45 and older accounted for a large proportion of all criminals.

5. Strip Search & Actions at Arrest

There were 6 behaviors at the time of arrest, namely 'concealment', 'resistance', 'insanity', 'being attacked' and 'cooperation'. Figure 6 shows the number of people with and without the corresponding behaviors.

We conducted six hypothesis tests for each of the behaviors at the time of arrest. Similar to the T-test for Sex vs. StripSearch, all three T-tests assumptions were checked and only normal distribution assumptions were violated.

H0: the mean of strip search for people have X aggressive action and people do not have X aggressive action are not different

H1: the mean of strip search for two groups are different

(Where X is one of Concealed, Combative, Resisted, Mental_inst, Assaulted, Cooperative)

With a pre-established alpha level of 0.05, we can conclude that there is a statistically significant difference in the probability of strip search between individuals with and without each behavior. The corresponding test statistics and p-values are as follows: (10.197, $p < 0.001$), (14.911, $p < 0.001$), (9.96, $p < 0.001$), (16.42, $p < 0.001$), (6.83, $p < 0.001$), and (-4.81, $p < 0.001$). Therefore, we can reject the null hypothesis for each of the six tests.

Figure 6: Distributions for 6 Actions at Arrest

Method

To determine if there was a statistically significant relationship between the dependent variable and independent variables, we conducted T-tests, Tukey's test, and Chi-squared tests. We selected the significant independent variables for further analysis and modeling.

To achieve our research objective, we utilized two main methods. For research question 1, we performed ANCOVA analysis to assess if the means of strip search probability for three or more groups were significantly different from each other. Prior to ANCOVA analysis, we conducted a power analysis to calculate the statistical power of our test, the required sample size, and effect size requirements to achieve sufficient statistical power. We aimed to ensure that the probability of committing a Type II error was acceptable from a decision-making perspective.

For research question 2, we employed logistic regression as our model since we had a binary dependent variable. Our target outcome was being strip searched or not, and the predictors were significant demographic variables selected during EDA. To evaluate model performance, we divided the data into training and testing sets, with 80% of the data used for training. We evaluated the model on both the training and testing sets, measuring accuracy scores and confusion matrices to check the estimated and predicted accuracy. Finally, we calculated and plotted the prediction interval by varying one independent variable while controlling for the rest of the independent variables.

The assumptions for each model and tests were checked prior to modeling. This enabled us to ascertain that the results from the tests and models were reliable and accurate.

Results and Findings

Research Question 1:

Will the probability of strip search be influenced by people's gender groups when arrested?

Power Analysis

Before conducting ANCOVA analysis, we did power analysis to help to determine the appropriate sample size needed to detect a statistically significant effect.

To determine the required sample size for each group to achieve a specific level of statistical power, we used the TTestIndPower() function and specified the desired power level as the input parameter. We also calculated the ratio of one sample group against the other based on the information available in our dataset.Next, we calculated the effect size of the explanatory variable using Cohen's D metric, which was found to be -0.07. This effect size is considered small according to Cohen's guidelines, as it is smaller than 0.2.

Using the calculated effect size and setting the desired level of statistical power at 80%, we then computed the required sample sizes for each group. The results showed that a sample size of 2099 was needed for the group with family size less than three, while a sample size of 8762 was needed for the group with family size greater than three. It is important to note that the actual sample sizes in our dataset were 12609 and 52634 for these respective groups. Also , power is calculated which is 1 and we plot the diagram in the following. Thus, our analysis indicates that the sample sizes in the dataset are sufficient to achieve the desired level of statistical power.

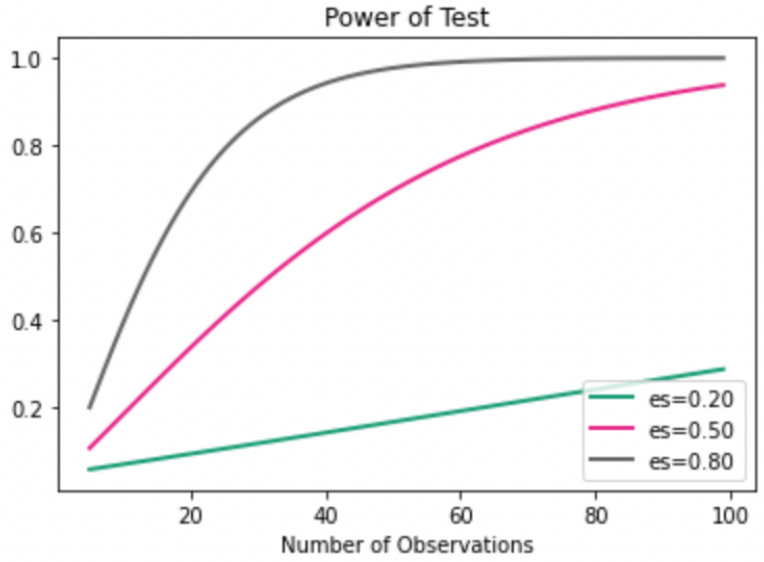

Figure 7: Power curve of the ANCOVA analysis

ANCOVA

We conduct an ANCOVA analysis to investigate whether there is a significant difference in stripe search probability between different gender groups while controlling their degree of actions.

H0: People from all age groups have the same strip search probability after considering their degree of actions.

	SS	Degree of Freedom	F-score	P-value
Age	19.499	3	62.30	< 0.001
DegreeOfAction	40.648	1	389.60	< 0.001
Residual	6806.502	65238		

Table 7: ANCOVA Result

By analyzing the result, we can have two interpretations:

Statistical interpretation Interpretation p-unc = “uncorrected p-value” for Age group is smaller than 0.05. We reject the null hypothesis that each of the age group results in the same strip search rate, even after controlling for the degree of action.This finding suggests that age is an important factor in determining the likelihood of being subjected to strip searches by police, and that individuals in certain age groups may be more likely to be searched than others.

Practical interpretation We hypothesized that age group has no relationship with the strip search rate. However, the results show that age group is firmly related to the strip search rate even controlling the degree of action.These findings have practical implications for law enforcement agencies and policymakers who may want to consider ways to reduce the use of strip searches and ensure that they are applied fairly and consistently across different age groups.

Logistic Regression

Research Question 2:

Will the probability of strip search be influenced by people's gender, age, race, and their actions when arrested?

We perform a logistic regression to examine the effects of degree of aggressive actions at arrest, age, sex, and race on the likelihood that criminals will be strip searched.

Formula for logistic model:
StripSearch ~ Degree Of Action + Aged 25 to 34 years + Aged 35 to 44 years + Aged 45 and older + Male + Black + Other + White

Variable	Description
Degree of Action	sum of 6 actions' indicator, ranged from 0 to 6
M	1 for Male, 0 for Female
Aged_25_to_34_years	1 when the person is aged 25 to 34, 0 otherwise
Aged_35_to_44_years	1 when the person is aged 35 to 44, 0 otherwise
Aged_45_and_older	1 when the person is aged 45 and order, 0 otherwise
Black	1 when the person's Race is Black, 0 otherwise
Other	1 when the person's Race is Other, 0 otherwise
White	1 when the person's Race is White, 0 otherwise

Table 8: Variable Description For Logistic Model

Before conducting logistic regression, there are a few model assumptions that were checked to ensure the prediction power.

a. Binary Response Variable: Strip Searched or Not Strip Searched
b. Linearity: It's hard to check before fitting the model as the target is a binary value, we check it later when we get a predicted probability. The linearity plot is shown in Figure 8
c. Independence of Errors: The errors and measurements are all independent with each other
d. Equal variance: Almost equal across all levels of the independent variables. The variance table of all groups can be found in Appendix Section.

e. No multicollinearity: all VIF < 5

Features	VIF Factor
DegreeOfAction	1.8229
Aged_25_to_34_years	2.2064
Aged_35_to_44_years	1.9664
Aged_45_and_older	1.9236
M	3.8871
Black	2.1006
Other	1.7795
White	2.9318

Table 9: VIF Table

f. No influential outliers: There are no influential outliers in the data.

g. Large sample size: the strip searched groups have relatively small sample size, this is one limitation of the model. The table of cunts can be found in the Appendix Section.

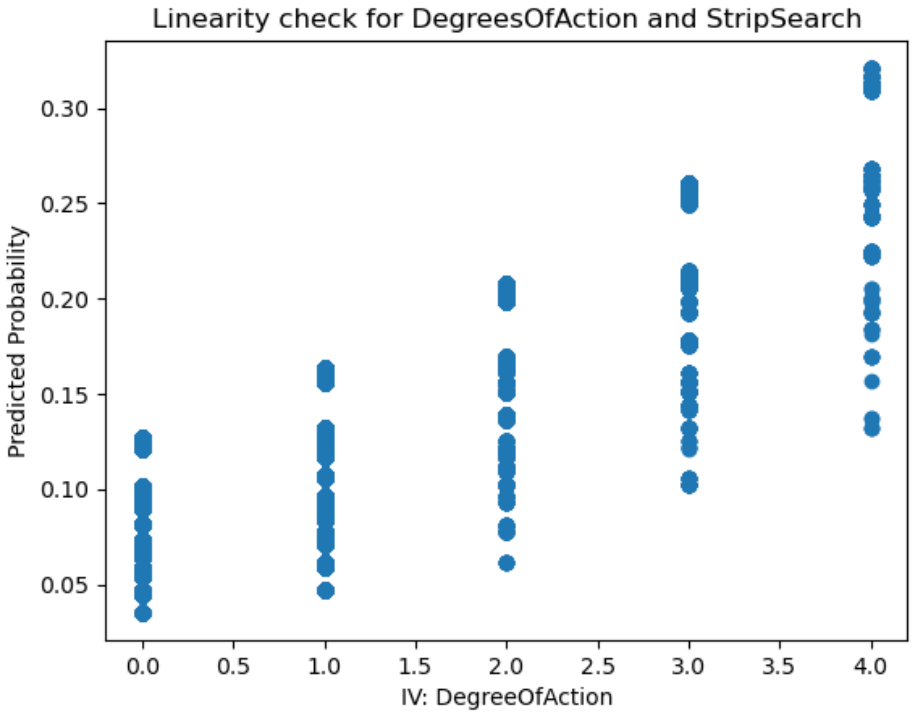

Figure 8: Linearity Check for Degree of Action and Strip Search

Logistic Model Result

	Coefficient	Standard Error	Z score	P-value	CI: [0.025 ,	0.975]
Intercept	-2.8584	0.064	-44.913	< 0.001	-2.983	-2.734
DegreeOfAction	0.2944	0.017	17.148	< 0.001	0.261	0.328
Aged_25_to_34_years	0.0339	0.038	0.898	0.369	-0.04	0.108
Aged_35_to_44_years	-0.005	0.04	-0.125	0.901	-0.084	0.074
Aged_45_and_older	-0.4525	0.045	-10.017	< 0.001	-0.541	-0.364
M	0.2526	0.036	6.964	< 0.001	0.182	0.324
Black	0.6267	0.054	11.597	< 0.001	0.521	0.733
Other	0.2937	0.059	4.979	< 0.001	0.178	0.409
White	0.645	0.052	12.362	< 0.001	0.543	0.7

Table 10: Logistic Model 1 Result

	Lower CI	Upper CI	OR
Intercept	0.050636	0.064984	0.057363
DegreeOfAction	1.297912	1.388267	1.342329
Aged_25_to_34_years	0.960736	1.113784	1.034434
Aged_35_to_44_years	0.919387	1.0768	0.994985
Aged_45_and_older	0.582111	0.69489	0.636006
Male	1.199063	1.3823	1.287426
Black	1.683357	2.080546	1.871444
Other	1.194924	1.505807	1.34139
White	1.720745	2.111258	1.90602

Table 11: Predicted Interval and Odds Ratio

From the result above, we found the DegreeOfAction, Aged 45 and order, Sex, Race are statistically significant on the likelihood of being strip searched.

When DegreeOfAction is 0 (not concealed, combative, resisted, mental_inst, assaulted, or not cooperative), Asian female criminal aged 24 and younger would have the log of the odds of being strip searched as -2.86, which means an odds of being strip searched is 0.05.
Increasing the degree of action was associated with an increased likelihood of being strip searched. For each additional degree of action increase for the criminal, the odds that the criminal will be strip searched increases by about 1.34 times, controlling for the other features.

When holding other features constant,

- Compared with criminals who are aged 24 or below, the log of the odds of being strip searched among those criminals who aged 45 or older are 0.45 lower. The odds of being strip searched is 0.63 times higher (1.57 times lower) for people aged 45 or older than people aged 24 or younger.
- Compared with female criminals, the log of the odds of being strip searched among male criminals are 0.25 higher. The odds of being strip searched is 1.28 times higher for male than females.
- Compared with Asian criminals, the log of the odds of being strip searched among Black criminals are 0.62 higher. The odds of being strip searched is 1.87 times higher for Black criminals than Asian criminals.
- Compared with Asian criminals, the log of the odds of being strip searched among Other race criminals are 0.29 higher. The odds of being strip searched is 1.34 times higher for other race criminals than Asian criminals.
- Compared with Asian criminals, the log of the odds of being strip searched among White criminals are 1.92 higher. The odds of being strip searched is 1.91 times higher for White criminals than Asian criminals.

Predictive Interval

To create a 2D map of the predictive interval, we selected the independent variable 'degree_of_action' as the variable on the x-axis while holding all other variables constant. To determine the values for the remaining variables, we utilized the groups that were most frequently strip searched based on our contingency heat map presented in Figure 5. Specifically, we selected white, male criminals aged 25 to 34 as our reference group.

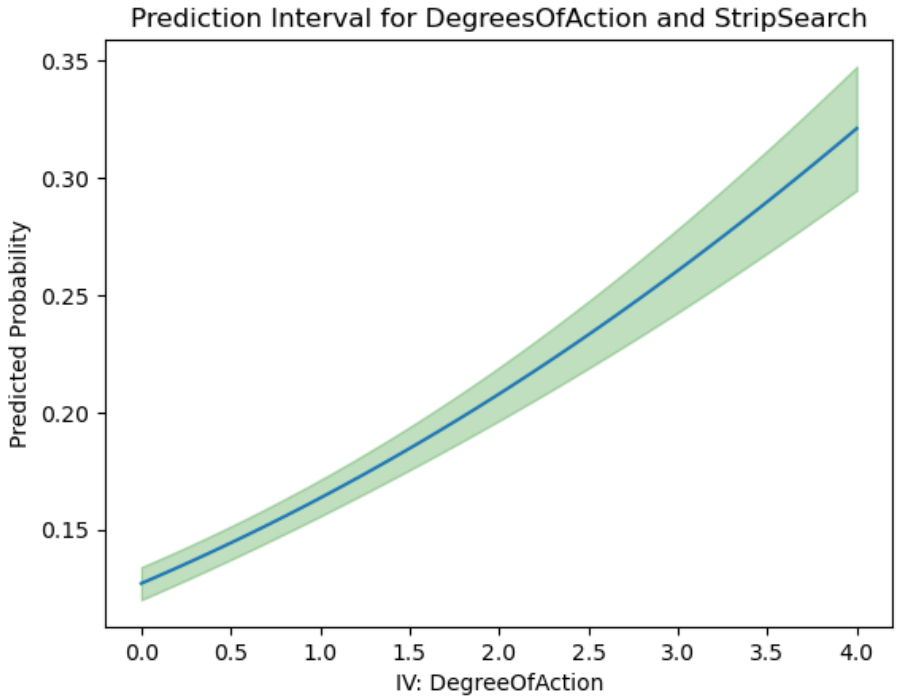

Figure 9: Prediction Interval for Degree of Action and Strip Search

As shown in Figure 9, the predictive interval plot reveals a positive relationship between the degree of action and the probability of being strip searched, which confirms the results of our previous model. This interval is a 95% prediction interval, as we chose an alpha value of 0.05. When the degree of action is 0, the predicted probability of being strip searched for white male criminals aged 25 to 34 is approximately 0.12. In contrast, when the degree of action is 4, the predicted probability of being strip searched is around 0.32, with a wider confidence interval. This indicates a higher level of uncertainty regarding the range of values the target could take when DOA is 4, with a lower bound of approximately 0.29 and an upper bound of about 0.35.

Model Peformance

Estimated accuracy (prediction on the training set) is 0.88 and the accuracy on the testing set is 0.88 as well.

True\Predicted	Not Searched	Searched
Not Searched	45962	0
Searched	6232	0

Table 12: Confusion Matrix on Training Set

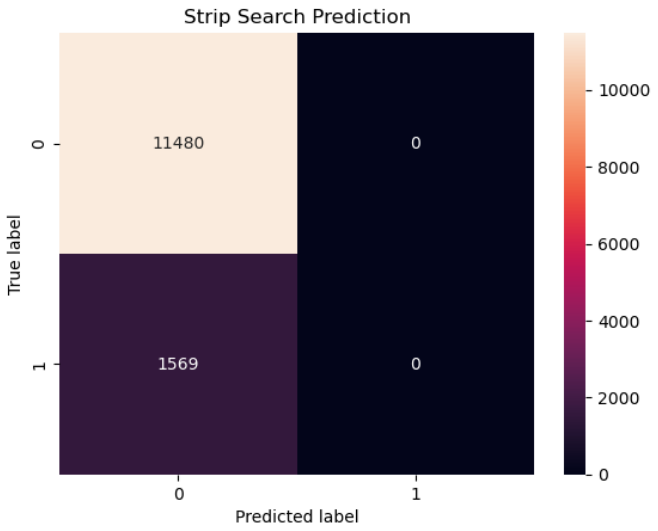

Figure 10: Confusion Matrix on Testing Set

The confusion matrices reveal that (1) all strip searched observations are classified as not strip searched. (2) The error rate is only 2.3%, indicating that the model is performing well in correctly identifying individuals who are not strip searched. The accuracy rate is also high, at 97.7%, due to the high proportion of negative cases in the dataset. (3) The precision and recall metrics are 0 (undefined), as there are no true positives in the predicted strip searched group.

As a result, although the model had a high accuracy rate, the model did not perform an accurate classification, as all strip-searched observations were classified as not strip-searched. Furthermore, we found that since the maximum predicted probability is only 0.32, the threshold we used (0.5) to classify and assign labels had classified all strip searched observations into the not strip searched group. This could indicate that our model is kind of underfitting. To address these issues, we explored two potential solutions to improve its performance:

(a) Explore the possibility of adjusting the threshold value used to classify the observations. Instead of only using accuracy score for model evaluation, we may also calculate the model precision, recall, and f-score.

(b) Add more variables and features to the model

For solution (a), as Figure 11 showed, accuracy drops when the threshold is far away from 0.32, while the f-score is relatively high and reaches its maximum when the threshold is 0.1. After testing different thresholds, we found that a value of 0.165 provided a good balance between accuracy and f-score. This threshold achieved a test accuracy of 0.855, with improved precision, recall, and f-score.

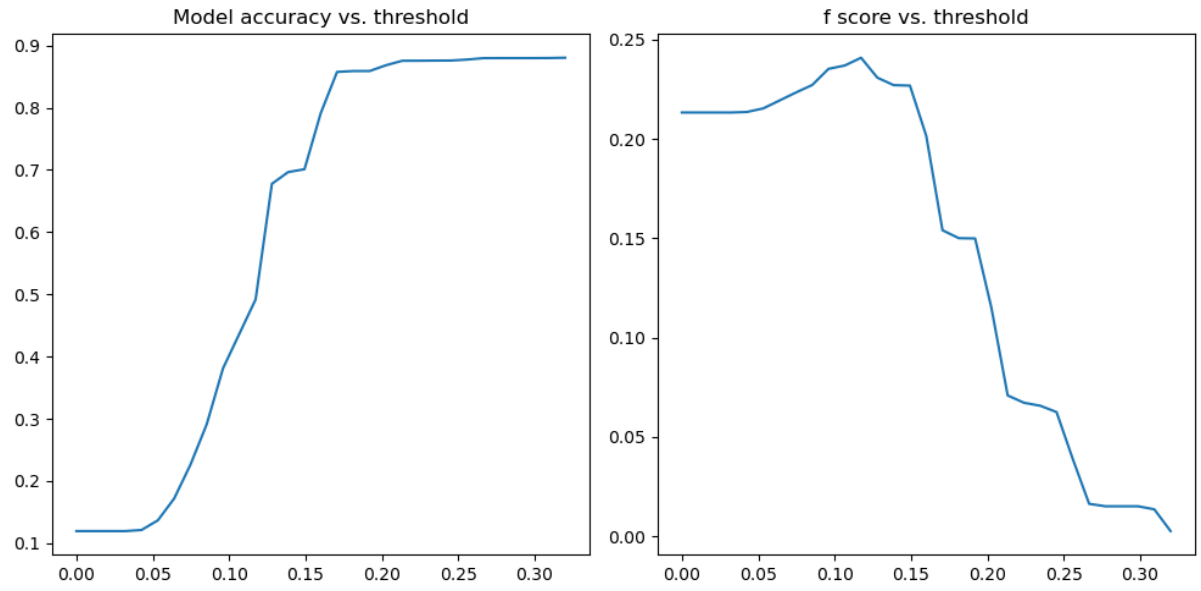

Figure 11: Line Plots for Accuracy and F-score vs. threshold

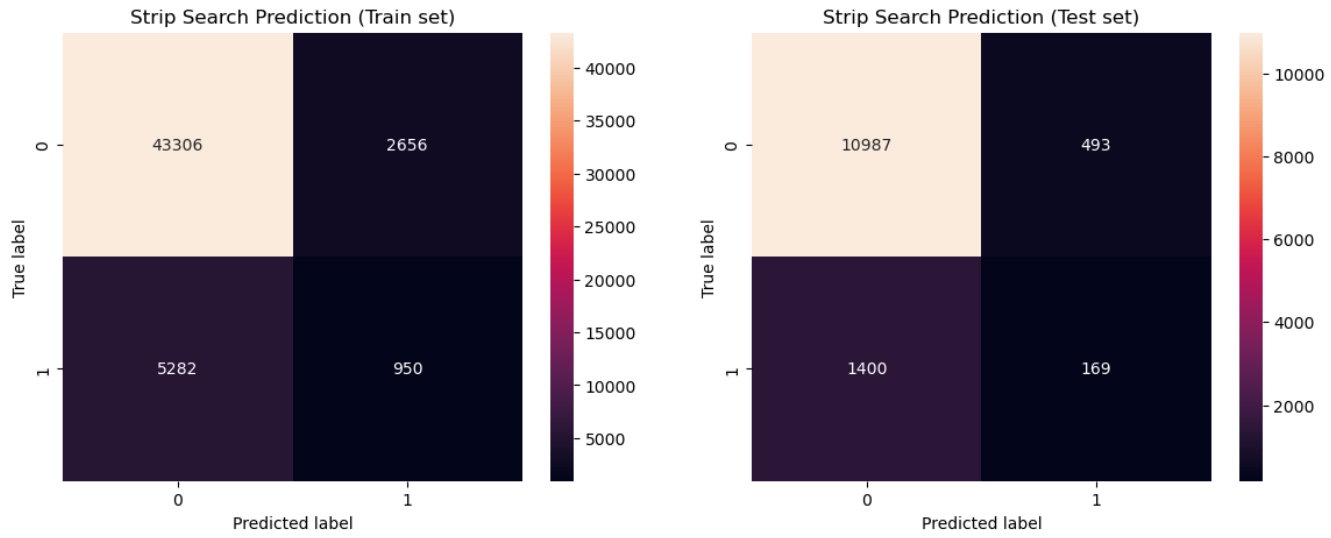

Figure 12: Confusion Matrices on Training and Testing Set

For solution (b), rather than using the Degree of Action variable, which sums the number of aggressive actions an individual exhibits when arrested, we used the original binary indicators for each action:

StripSearch ~ Concealed + Combative + Resisted + Mental_inst + Assaulted + Cooperative + Aged 25 to 34 years + Aged 35 to 44 years + Aged 45 and older + Male + Black + Other + White

Result for the new logistic model:

	Coefficient	Standard Error	Z score	P-value	CI: [0.025 , 0.975]	
Intercept	-2.7394	0.064	-42.573	< 0.001	-2.866	-2.613
Concealed	1.4231	0.142	10.011	< 0.001	1.144	1.702
Combative	0.5468	0.059	9.301	< 0.001	0.432	0.662
Resisted	0.2683	0.063	4.271	< 0.001	0.145	0.391
Mental_inst	0.8232	0.059	13.873	< 0.001	0.707	0.939
Assaulted	0.3803	0.134	2.847	0.004	0.118	0.642
Cooperative	0.0158	0.029	0.551	0.582	-0.040	0.072
Aged 25- 34	0.0282	0.038	0.744	0.457	-0.046	0.103
Aged 35-44	-0.0096	0.041	-0.237	0.813	-0.089	0.070
Aged 45 or older	-0.4570	0.045	-10.069	< 0.001	-0.546	-0.368
Male	0.2512	0.036	6.887	< 0.001	0.180	0.323
Black	0.6258	0.054	11.547	< 0.001	0.520	0.732
Other	0.2968	0.059	5.018	< 0.001	0.181	0.413
White	0.6452	0.052	12.336	< 0.001	0.543	0.748

Table 13: Result for Logistic Model 2

As shown in Table 13, all actions except for "Cooperative" were found to have an effect on the probability of being strip-searched. The new model achieved an accuracy of 0.847 on the testing set, with a positive TP score of 228, indicating that more positive (strip-searched) cases were correctly identified.

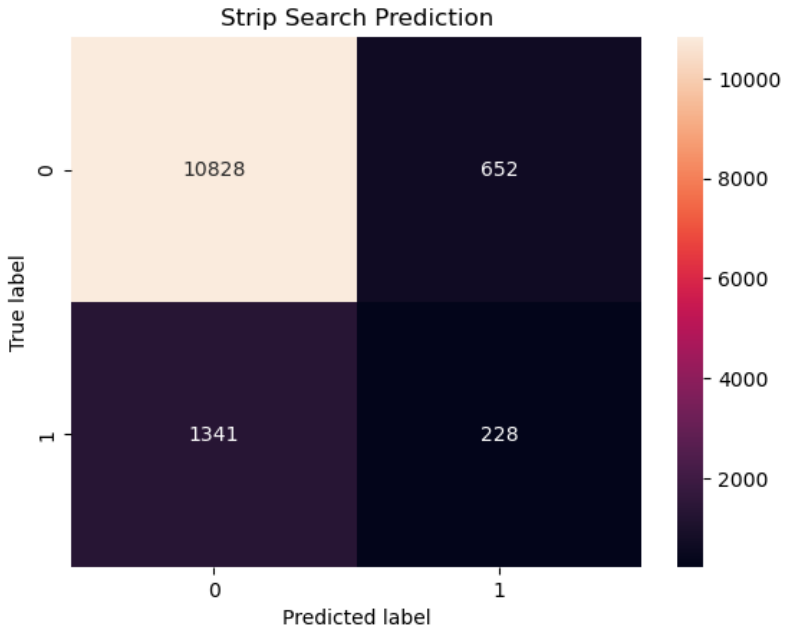

Figure 13: Confusion Matrix on Testing Set

By implementing these solutions, we were able to improve the model's performance and achieve a more accurate classification of strip-searched individuals.

Discussion

Based on the results obtained from T-tests, Tukey's Test, Chi-squared tests, ANCOVA, and Logistic Regression, we can now answer our two research questions.

Question 1:

The ANCOVA analysis suggests that there is a significant relationship between age group and strip search probability, even after controlling for the effect of degree of action. This finding is supported by the uncorrected p-value, which is less than 0.05, indicating that the probability of obtaining these results by chance is very low. In addition, our power analysis suggests that our study had a very high statistical power of 1.000, meaning that it was very likely to detect a true difference between groups, even with a relatively small effect size. It suggests that our results are reliable and can be trusted.

After all, These results suggest that age is an important factor in determining the likelihood of being subjected to strip searches by police, even when controlling for the effect of degree of action. This finding has practical implications for law enforcement agencies and policymakers, as it highlights the need to ensure that strip searches are applied fairly and consistently across different age groups.

Question 2:

The probability of strip search was influenced by people's gender, age, race, and their actions when arrested. More specifically, increasing the degree of action was associated with an increased likelihood of being strip searched. The odds of being strip searched is higher for criminals aged 24 or younger than those aged 45 or older, higher for male than females, higher for Black criminals than Asian criminals, higher for other race criminals than Asian criminals, higher for White criminals than Asian criminals.

We are 95% confident that: predicted mean strip search probability for male is between 1.2 and 1.38 times higher than females; predicted mean strip search probability for White criminals is between 1.72 and 2.11 times higher than Asian Criminals; predicted mean strip search probability for people aged 24 or younger is between 1.44 to 1.72 times higher than people aged 45 or older.

Limitation

In our dataset, some individuals have been arrested multiple times, which poses a challenge to analyzing associated measurements. In our project, we assumed each arrest as an independent individual case, resulting in non-independent and related data and findings. For instance, the strip search probability may increase when an individual has been arrested multiple times. In future analyses, we plan to incorporate a new variable in our model to record the number of previous arrests.

We used the binary variable StripSearch as the outcome for our ANCOVA test to approximate a search probability. However, this may lead to estimated values or confidence intervals that do not make sense, such as negative probabilities. Additionally, we created a

Degree of Action variable to approximate a continuous variable, even though it is actually a discrete variable. This could result in inappropriate and misleading ANCOVA results.

Furthermore, the logistic model's performance is hindered by the imbalanced dataset, where only 12% of individuals are strip searched. Consequently, the model excels in identifying individuals who are not strip searched but performs poorly in identifying strip-searched cases. We recommend adjusting the threshold value used to classify and assign labels for observations, incorporating more features to the model to avoid underfitting. And in the future, we can utilize weighted logistic models to address the imbalanced distribution, employ resampling methods to simulate more strip-searched cases, and implement stratification strategies when splitting the training and testing sets.

Additionally, due to the imbalance in the dataset, the distribution within each variable group is highly skewed. This makes the normality assumption for T-tests strongly violated, and the T-test may not be suitable for analyzing the relationships between two categorical variables. To address this problem, we conducted a Chi-squared test, which satisfied all test assumptions and provided reliable and powerful results.

Conclusion

In summary, our analysis of the arrest data revealed clear evidence of bias in the strip search practices of law enforcement officers. Specifically, our logistic regression analysis found that being male, Black, or White, aged 25 to 34 were significant predictors of being strip searched, even after controlling for other variables such as the degree of aggression displayed during the arrest. Additionally, we found that the probability of being strip searched decreased with increasing age. These findings suggest that there are significant disparities in the way strip searches are conducted based on an individual's demographic characteristics. Such biased practices not only lead to a violation of individual rights but also undermine the public's trust in law enforcement. Our findings highlight the importance of continued efforts to reduce bias and promote fairness in law enforcement practices, particularly in the area of strip searches. Policies such as the mandatory recording of all strip searches and increased training on implicit bias for law enforcement officers could be implemented to address these issues.

Appendix

Sex	Age_group	Perceived_Race	Variance
F	Aged_24_years_and_under	Asian	0.04676953
F	Aged_24_years_and_under	Black	0.08499467
F	Aged_24_years_and_under	Other	0.08415817
F	Aged_24_years_and_under	White	0.08310635
F	Aged_25_to_34_years	Asian	0.04182487
F	Aged_25_to_34_years	Black	0.07401952
F	Aged_25_to_34_years	Other	0.09525004
F	Aged_25_to_34_years	White	0.13824778
F	Aged_35_to_44_years	Asian	0.04030692
F	Aged_35_to_44_years	Black	0.06771739
F	Aged_35_to_44_years	Other	0.0807512
F	Aged_35_to_44_years	White	0.1240884
F	Aged_45_and_older	Asian	0.02977642
F	Aged_45_and_older	Black	0.08307954
F	Aged_45_and_older	Other	0.05577346
F	Aged_45_and_older	White	0.06734494
M	Aged_24_years_and_under	Asian	0.08883254
M	Aged_24_years_and_under	Black	0.14214005
M	Aged_24_years_and_under	Other	0.09877132
M	Aged_24_years_and_under	White	0.10280471
M	Aged_25_to_34_years	Asian	0.07218336
M	Aged_25_to_34_years	Black	0.1292106
M	Aged_25_to_34_years	Other	0.09676217

M	Aged_25_to_34_years	White	0.12805768
M	Aged_35_to_44_years	Asian	0.08795065
M	Aged_35_to_44_years	Black	0.12138234
M	Aged_35_to_44_years	Other	0.10227125
M	Aged_35_to_44_years	White	0.12449413
M	Aged_45_and_older	Asian	0.04879139
M	Aged_45_and_older	Black	0.10261178
M	Aged_45_and_older	Other	0.06444017
M	Aged_45_and_older	White	0.08956655

Table for Equal Variance Check

Sex	Age_group	Perceived_Race	Not Searched	Searched
F	Aged_24_years_and_under	Asian	291	15
F	Aged_24_years_and_under	Black	919	95
F	Aged_24_years_and_under	Other	451	46
F	Aged_24_years_and_under	White	895	90
F	Aged_25_to_34_years	Asian	329	15
F	Aged_25_to_34_years	Black	972	85
F	Aged_25_to_34_years	Other	705	84
F	Aged_25_to_34_years	White	1703	338
F	Aged_35_to_44_years	Asian	297	13
F	Aged_35_to_44_years	Black	534	42
F	Aged_35_to_44_years	Other	464	45
F	Aged_35_to_44_years	White	1444	245
F	Aged_45_and_older	Asian	285	9
F	Aged_45_and_older	Black	329	33

F	Aged_45_and_older	Other	366	23
F	Aged_45_and_older	White	1342	105
M	Aged_24_years_and_under	Asian	1355	148
M	Aged_24_years_and_under	Black	3381	700
M	Aged_24_years_and_under	Other	1865	233
M	Aged_24_years_and_under	White	2295	302
M	Aged_25_to_34_years	Asian	1837	156
M	Aged_25_to_34_years	Black	4471	804
M	Aged_25_to_34_years	Other	2851	347
M	Aged_25_to_34_years	White	5306	942
M	Aged_35_to_44_years	Asian	1446	156
M	Aged_35_to_44_years	Black	2619	431
M	Aged_35_to_44_years	Other	2150	281
M	Aged_35_to_44_years	White	5189	885
M	Aged_45_and_older	Asian	1587	86
M	Aged_45_and_older	Black	1859	244
M	Aged_45_and_older	Other	1937	144
M	Aged_45_and_older	White	5968	659

Contingency Table for Sufficient Sample Check